

FEATURES	source
550. aggaacagcttcagagatctctgatacttctcgggaacccattttaaactgataat	609
1 aggacacacattctcagatgcttgcattctttaaactgggaacccattttaaactgataat	60
610 tgcacaatgttcagattctcctctcctgtaatgcagaatgtgcacaagtcttcggaatttcg	669
61 ttcacacaaattgttcagatttctcctctcctgtaatgcagaatgtgcacaagtcttcggaatttgc	120
670 tctgttcacagcatcagacagaaatgtgtccagaatccttttgacaacactgcgtccagatggg	729
121 tctgttccacagcatcagacagaaatgtgtccagaatccttttgacaacactgcgtccagatggg	180
730 gcaagactctcaatgtttcttcaatcacagagaaaaaattctattctgtgaaatagataag	789
181 gcagaaactcttattgttttctttaaactacatgagaaaaaattctattgtgaaatagataag	240
790 agtaactggagctctcaatgtgtctgtatcagtaaaagggcaaaatcacagatgtgagacaga	849
241 aatfactggagctctcaatgtgtctgtatcagtaaaagggcaaaatcacagatgtgagacaga	300
850 ttgaaaagggaagaa--caatttcacatcctgtatcagatgttctcctaagaagggcagc	907
301 ttgaaaagggaagaaacatttttcattgcagatcgtgtaattgatactttaaataaagagcagc	360
908 -tcaagacatcacag 922	
361 ttccaagacatcacag 376	
RESULT 5	
AA897178/c	238 bp mRNA linear EST 04-JAN-1998
LOCUS	
DEFINITION	am9e08.s1 Soares_NFL.T.GRC_S1 Homo sapiens cDNA clone
IMAGE	1466342 3', mRNA sequence.
ACCESSION	AA897178
VERSION	AA897178.1 GI:3033798
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE	1 (bases 1 to 238)
COMMENT	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap . National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index Unpublished (1997) Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nih.gov This clone is available royalty-free through LINDL; contact the IMAGE Consortium (info@image.lindl.gov) for further information. Insert length: 847 Std Error: 0.00 Seq primer: -40m13 fwd. Err from Amersham High quality sequence stop: 132. Location/Qualifiers 1. 238

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1466342"
/clone_lib="Scates_NFL_T_CRC_S1"
/lab_host="DH10B"
/notes="Organ: pooled; Vector: p7T73D-Pac (Pharmacia) with
a modified polylinker; Site1: Not I; Site 2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung Nbhl19A, testis NHT, and B-cell
NCL-GAP-GCB1) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682633-687229,
726408-728711, and 729096-733139. Subtraction by Bento
Scates and M. Fatima Donaldso"

```

BASE COUNT	71 a	53 c	45 g	69 t
Query Match	16.1%	Score 226;	DB 9;	Length 238;
Best Local Similarity	99.6%	Pred. No. 4.5e-47;		
Matches 237; Conservative	0;	Mismatches 0;	Indels 1;	Gaps 1;
Oy 1156	actggcgtccatttcacatgtaatggtccatttgcacgtgatttcagccagaagaactg	1215		
Db 238	actggcgtccatttcacatgtaatggtccatttgcacgtgatttcagccagaagaactg	179		
Oy 1216	tggcgtgataataatggaatggtgtgttcgaaagtgcctgaaataacagtttaagcagca	1275		
Db 178	tggcgtgataataatggaatggtgtgttcgaaagtgcctgaaataacagtttaagcagca	119		
Oy 1276	ttaatagctcccgcaatg-tgaattgctaataacagtttaaaaaacggatattccctggt	1334		
Db 118	ttaatagctcccgcaatg-tgaattgctaataacagtttaaaaaacggatattccctggt	59		
Oy 1335	tctgtccacaattatcagttgtttacaagaacccctattcatgtgttaagaagacact	1392		
Db 58	tctgtccacaattatcagttgtttacaagaacccctattcatgtgttaagaagacact	1		
RESULT 6				
AA776169	209 bp	mRNA	linear	EST 05-FEB-1998
LOCUS	ae00c02.s1	Stratagene	schizo brain S11	Homo sapiens cDNA clone
DEFINITION	IMAGE:970466 3', mRNA sequence.			
ACCESSION	AA776169			
VERSION	AA776169.1	GI:2835503		
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
REFERENCE	1 (bases 1 to 209)			
AUTHORS	Hillier,L., Allen,M., Bowles,L., Dubuque,T., Gelsel,G., Jost,S.,			
	Krizman,D., Kucaba,T., Lacey,M., Le,N., Lennon,G., Merri,M., Martin			
	,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theisling,B.,			
	White,Y., Wylie,T., Waterston,R. and Wilson,R.			
TITLE	WashU-NCI human EST project			
COMMENT	Unpublished (1997)			
	Contact: Wilson RK			
	Washington University School of Medicine			
	4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108			
	Tel: 314 286 1800			
	Fax: 314 286 1810			
	Email: est@wustl.edu			
	This clone is available royalty-free through LNL; contact the			
	IMAGE Consortium (info@image.llnl.gov) for further information.			
	Possible reversed clone: polyT not found			
	Seq primer: -40ml3 fwd. 5' from Amersham			
	High quality sequence stop: 206.			
FEATURES	Location/Qualifiers			
Source	1..209			
	/organism="Homo sapiens"			
	/db_xref="taxon:9606"			
	/clone="IMAGE:970466"			
	/clone_idb="Stratagene schizo brain S11"			
	/sex="male"			
	/tissue-type="Schizophrenic brain S-11 frontal lobe"			
	/dev-stage="34 years old"			
	/lab_host="SOLR (kanamycin resistant)"			
	/note="Vector: Bluescript SK-; Site, 1: EcoRI; Library			
	constructed from S-11 frontal lobe, male, 34 years old,			
	50% caucasian, 50% Aleutian. Schizophrenic suicide.			
	Random primed into EcoRI site of ZAP II Vector. Mass			
	excised. Custom library. Avg insert length 1.4kb.			
	Material obtained by Johnston N., Torrey, E.F., Volken R.,			
	and the Stanley Neuropathology Consortium - Analysis of			
	RNAs from the Brains of Individuals with Psychiatric			

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 22, 2002, 05:31:11 ; Search time 3619.39 Seconds

(without alignments)
5224.426 Million cell updates/sec

Title: US-09-693-205-7_COPY_5300_6700

Perfect score: 1401
Sequence: 1 gtagcagtaaacactagagc.....aaagacacttaagaagt 1401

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Archived: 13736207 segs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estbta:*
2: em_estbta:*
3: em_estin:*
4: em_estinu:*
5: em_estov:*
6: em_estopl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_hic:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pin:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	752.2	53.7	1083	BM476887	BM476887 AGENCOURT
2	673.4	48.1	795	R17106	R17106 EST20108 C1
3	458.4	32.7	1010	BM472954	BM472954 AGENCOURT
4	345.2	24.6	376	AL596565	AL596565 DFF2451L
5	226	16.1	238	AA897178	AA897178 am09608.s
6	203.4	14.5	209	AA776169	AA776169 ae80602.s
7	142.4	10.2	148	AA776670	AA776670 ae80602.s
8	134.6	9.6	499	A2177954	A2177954 SP_0148.B
9	114.8	8.2	505	AW295072	AW295072 UI-H-BW0-
10	107.8	7.7	434	AW470182	AW470182 xw61a11.x
11	87.6	6.3	1067	CNS02W80	AL216729 Tetradon
12	87.2	6.2	586	BF650306	BF650306 NF087D1E
13	85	6.1	918	A2136287	BF650306 Tetradon
14	84	6.0	807	A2185191	A2136287 SP_0169.B
15	80	5.7	431	A2182120	A2185191 SP_1004.B
16	79.6	5.7	331	A1685137	A2182120 SP_0188.A
17	67.2	4.8	821	BH601255	A1685137 wa75a03.x
					BH601255 BOGYW28TR

18	61.6	4.4	643	9	BB485245	BB485245
19	58.4	4.2	284	9	BB254184	BB254184
20	58.2	4.2	445	10	BF646344	BF646344 NF074E04E
21	57.8	4.1	616	9	A2963183	A2963183 2M0232F17
22	55.2	3.9	234	9	BB254737	BB254737
23	51.4	3.7	721	12	AQ254586	AQ254586 CPG0711B
24	48.4	3.5	1101	12	CNS0039G	AL063921 Drosophila
25	47.8	3.4	1069	12	CNS05ESA	AL334039 Tetradon
26	47.2	3.4	936	12	CNS040NC	AL212803 Tetradon
27	46	3.3	959	12	CNS040NC	AL302817 Tetradon
28	44.8	3.2	1101	12	CNS0006J	AL062049 Drosophila
29	44.4	3.2	555	12	AQ310011	AQ310011 CITBI-El-
30	44.4	3.2	560	12	AQ310013	AQ310013 CITBI-El-
31	43.8	3.1	937	12	CNS040H	AL291194 Tetradon
32	42.6	3.0	380	9	AM517860	AM517860 xub1f12.x
33	42.6	3.0	407	9	AM294501	AM294501 UI-H-B12-
34	42.6	3.0	548	10	BM274569	BM274569 PFES0aa4
35	42.6	3.0	609	10	BE440138	BE440138 HTM1-946A
36	42.4	3.0	670	12	CNS014PQ	AQ308275 CITBI-El-
37	42.2	3.0	987	12	CNS014PQ	AL104456 Drosophila
38	42	3.0	979	12	CNS00D9H	AL060395 Drosophila
39	41.8	3.0	536	12	AQ757450	AQ757450 HS_5466_B
40	41.8	3.0	782	9	AL552607	AL552607
41	41.6	3.0	496	12	BM479411	BM479411 BOGX20TF
42	41.6	3.0	524	12	BH600863	BH600863 BOHOK30TF
43	41.6	3.0	547	12	BH502803	BH502803 BGDPI15TR
44	41.6	3.0	611	12	BH436907	BH436907 BOHAI019TR
45	41.6	3.0	616	12	BH463467	BH463467 BOHNA40TF

ALIGNMENTS

RESULT 1
BM476887
LOCUS
DEFINITION
AGENCOURT_6481789 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:5555441
5', mRNA sequence.
ACCESSION
BM476887.1 GI:18525929
VERSION
EST.
KEYWORDS
SOURCE
ORGANISM
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 1083)
NIH-MGC <http://mhc.nci.nih.gov/>.
AUTHORS
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished (1999)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLMA12275 row: 3 column: 18
High quality sequence stop: 696.
Location/Qualifiers
1. 1083
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5555441"
/clone.lib="NIH_MGC_71"
/tissue.type="leiomysarcoma"
/lab.host="DH10B (phage-resistant)"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site:1: NotI;
Site:2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2.1 kb."
BASE COUNT
334 a 219 c 212 g 318 t
ORIGIN

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 22, 2002, 06:53:41 ; Search time 3530.57 Seconds

(without alignments)
3514.853 Million cell updates/sec

Title: US-09-693-205-7_COPY_12200_12792

Perfect score: 593

Sequence: 1 acatctatgtttacagctc.....accctccattttatgctt 593

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Matched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba:*
2: gb_hlg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_ph:*
7: gb_pl:*
8: gb_pr:*
9: gb_ro:*
10: gb_sts:*
11: gb_sy:*
12: gb_un:*
13: gb_vl:*
14: em_ba:*
15: em_fun:*
16: em_hum:*
17: em_mu:*
18: em_mu:*
19: em_mu:*
20: em_mu:*
21: em_mu:*
22: em_mu:*
23: em_mu:*
24: em_mu:*
25: em_mu:*
26: em_mu:*
27: em_mu:*
28: em_mu:*
29: em_mu:*
30: em_hlg_hum:*
31: em_hlg_inv:*
32: em_hlg_other:*
33: em_hlg_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
------------	-------	-------------	--------	-------	-------------

1	593	100.0	4318	9	AB018273	AB018273 Homo sapi
2	593	100.0	12793	6	AX119931	AX119931 Sequence
3	593	100.0	12793	9	AF193556	AF193556 Homo sapi
4	593	100.0	92693	9	AL157766	AL157766 Human DNA
5	593	100.0	99819	2	AC079761	AC079761 Homo sapi
6	426.6	71.9	3289	9	AB056815	AB056815 Macaca fa
7	327.4	55.2	418	11	G36555	G36555 SHGC-53325
8	61	10.3	169794	2	AC004688	AC004688 Plasmodu
9	61	10.3	196149	2	AC004709	AC004709 Plasmodu
10	59.6	10.1	4601	3	DM031584	DM031584 Drosophila
11	59.6	10.1	19517	3	AE001426	AE001426 Plasmodu
12	57.4	9.7	12029	3	AE001426	AE001426 Plasmodu
13	57.2	9.6	60604	2	AC023466	AC023466 Homo sapi
14	57	9.6	183813	2	AC012204	AC012204 Homo sapi
15	57	9.6	197225	2	AC093835	AC093835 Homo sapi
16	56.8	9.6	110000	2	PFMAL4P1_2	PFMAL4P1_2 Continuation (3 of
17	56.6	9.5	11691	6	AX347143	AX347143 Sequence
18	56.6	9.5	349980	6	AX344564	AX344564 Sequence
19	55.4	9.3	124635	6	AP000593	AP000593 Homo sapi
20	55.2	9.3	1815	1	AF267220	AF267220 Candidatu
21	55.2	9.3	162075	9	HS127D3	HS127D3 Homo sapi
22	55	9.3	1141	6	AX083744	AX083744 Sequence
23	55	9.3	13038	6	AX346176	AX346176 Sequence
24	54.8	9.2	145598	6	AC008132	AC008132 Homo sapi
25	54.8	9.2	15387	6	AX345086	AX345086 Sequence
26	54.6	9.2	4611	3	PEA132006	PEA132006 Plasmodu
27	54.4	9.2	7918	8	AF325123	AF325123 Arabidops
28	54.4	9.2	13574	6	AX346218	AX346218 Sequence
29	54.4	9.2	62268	9	HS1178121	HS1178121 Human DNA
30	54.4	9.2	141748	8	AC009323	AC009323 Arabidops
31	54.4	9.2	187048	2	AC024921	AC024921 Homo sapi
32	54	9.1	349980	6	AX344558	AX344558 Sequence
33	54	9.1	349980	6	AX344559	AX344559 Sequence
34	53.8	9.1	863	11	CNS06EVO	CNS06EVO T7 end of
35	53.8	9.1	189790	2	AC107420	AC107420 Homo sapi
36	53.4	9.0	186625	9	AC006487	AC006487 Homo sapi
37	53	8.9	141469	9	AL139811	AL139811 Human DNA
38	53	8.9	161384	2	AC083801	AC083801 Homo sapi
39	52.8	8.9	17848	6	AX277865	AX277865 Sequence
40	52.8	8.9	17848	6	AX323550	AX323550 Sequence
41	52.8	8.9	17848	6	AX348363	AX348363 Sequence
42	52.8	8.9	19087	6	AX345694	AX345694 Sequence
43	52.8	8.9	17835	2	PFMAL13P2_3	PFMAL13P2_3 Continuation (4 of
44	52.6	8.9	150754	9	AC023491	AC023491 Homo sapi
45	52.6	8.9	175358	9	AC007981	AC007981 Homo sapi

ALIGNMENTS

RESULT 1	AB018273	4318 bp	mrna	linear	PRI 16-JUN-1999
LOCUS	AB018273				
DEFINITION	Homo sapiens mRNA for KIRA0730 protein, partial cds.				
ACCESSION	AB018273				
VERSION	AB018273.1	GI:3862180			
KEYWORDS	Homo sapiens adult male brain cDNA to mRNA, clone_11b:pb1uescriptII SK plus clone:hk03632.				
SOURCE	Homo sapiens				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.				
AUTHORS	Nagase, T., Ishikawa, K., Suyama, M., Kikuno, R., Miyajima, N., Tanaka, A., Kotani, H., Nomura, N. and Ohara, O.				
TITLE	Prediction of the coding sequences of unidentified human genes. XI. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro				
JOURNAL	DNA Res. 5 (5), 277-286 (1998)				
MEDLINE	99087487				
REFERENCE	2 (bases 1 to 4318)				
AUTHORS	Ohara, O., Suyama, M., Nagase, T., Ishikawa, K. and Kikuno, R.				
TITLE	Direct Submission				

THIS PAGE BLANK (USPTO)

JOURNAL Submitted (08-Oct-1998) Osamu Ohara, Kazusa DNA Research Institute, Laboratory of DNA Technology, Yana 1532-3, Kisarazu, Chiba 292-0812, Japan (E-mail: odanainfo@kazusa.or.jp, Tel:+81-438-52-3913, Fax:+81-438-52-3914)

FEATURES
source location/Qualifiers

1. 4318
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="hK03632"
/sex="male"
/issue_type="brain"
/clone_lib="pbluescriptII SK plus"
/dev_stage="adult"
1. 3015
/gene="KIAA0730"
1. 3015
/gene="KIAA0730"
/codon_start=1
/product="KIAA0730 protein"
/protein_id="BAA34450.1"
/db_xref="GI:3882181"
/translation="VTFTSWTEPLRNIGLKYLISQOOLGFAKESVYANENNSKE
TLQNTVDILHITFOBRMDLSGNFLKELSLIPKCPPEAPAEIRFHOYOVNGTL
PLIKNGAQNPKFKQCDVLOLMTSCPLPEKAPLSIKEGSDLGQEOLEOVN
MLNVLDPDLKVINNACNICNITTLDEEMVTRAKVLISYEFLSAEKREFFOLG
VAFVVEDGMLKLEPEEYINLEYESDFPYLYKLELGTGHOLEKHLGTDITSTG
OYEVLSRIEFKSEKGLDQPNMRYKRVSLFSLRQNDVSDLENVADLALYL
PSODGRVYSSILVDDAPHYKSRIOGNIGVMDVLSQYLGKDHGFFTKLMLEPO
KIRPLSLISLEQDEETPKVQCFALCSLGRLOLSSDQPTGLIRIKHENDN
AFLANEKAIRLCKALREGKLVSCFEKLTQTLRVGNFPIPHSRSEFAFLRFGNV
ILLYIOHSDSKDINFLAMLKSATMDLISDTSYLMLCNDIYRIGELDISGV
KDSSEPSKLELPMGTPPIPAIHITLMDPMNVFPGYGVLYVDAEGDLYGSDP
TYTVAIYOEVEDADNSPFLKITYQIDIGYSEKIVSLDLYKPSREESOSDS
ASTSPTEELPGLRSTPLPFGSESHKTSKSHQSPKTKVNSLPKLEKTSYVE
QAWKLPESERRKTIIRLYLKKHPDKPENHDYANEFKHLQNEINLEQATLDQAD
RSKRFTFSASRFQSDKYSFQRFYTSNQEATSHKSEKQONKEKCPSPSAQYTSOR
FVPPTEKSVGNFVARWLRQANRANFSANRANFSAANRANFCFKCYLSTKLALIAAD
YAVRGSKSDVPTALAKILEYLSOLELITNDVITLNAVYDLSLTRYDILPFPQI
PMDRTSEVAMVETCAIITIKLENFMQKV"

BASE COUNT 1395 a 734 c 847 g 1342 t
ORIGIN

Query Match 100.0%; Score 593; DB 9; Length 4318;
Best Local Similarity 100.0%; Pred. No. 4,7e-92;
Matches 593; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 acatctatgttaccaggtccctgttgaagaagatgaacggaacaaactaaatgtt 60
|||||
Db 3650 ACATCTTAGTTTACAGGCTTCTGTTGATGAAGATAGCAAGCAAACTAAATGTT 3709
|||||
Qy 61 ggcagttcttattaccagttgttgaatgttcttgcgaacgtctgcgaagaacatt 120
|||||
Db 3710 GGCAGTTCTTATTACAGGTTGATGATGTTCTGCAAACTCTGCGCAAGCAACATT 3769
|||||
Qy 121 tattaactgttagaacaactgtcttattgttgtgttacaatcttccacaatgttata 180
|||||
Db 3770 TATTAACTGTTAGAACACTGCTTATGTTGTGTGTACATATTTCACAAATGTTATA 3829
|||||
Qy 181 attatatagttgttgaacagatgcaactcttctgttcttcaaaagtctgcagtttaa 240
|||||
Db 3830 ATTATATAGTGTGTGAACAGATGCAATCTTTGTTGTCTGAAGGCTCTGCAGTTAA 3889
|||||
Qy 241 aaaaaaaacaaactcttcttcaatagtgaatgttgcagttcttcaacttaaaac 300
|||||
Db 3890 AAAAAAAACAACTTTCTTCAATATGCGATGTAGGGGATTTTAACTTAAAAAC 3949
|||||
Qy 301 atcaaaaatgttaaaatcaatctgttaactagtagttataactacagctataattc 360
|||||
Db 3950 ATCAAAAATGTAAATCACTTTGTGTATGTAGTAGTTTAAATATCGGCTTATATTTC 4009
|||||
Qy 361 cccatgaatgacagactgacattaatcatgttctgcacatgctcttcaactt 420
|||||

Db 4010 CCCATGATGATGCAAGTACATTTAAATTCAGTGTGTCTGCCCATGCTTCTTACTTT 4069
|||||
Qy 421 aacatattcttcttgaagaatgaaaggtaagatgaatgaattatataaagtactg 480
|||||
Db 4070 AACATATTTCTTTGCGAATGTAAAGGTAAAGATTAATAGTTTATTAAGTGTACTGG 4129
|||||
Qy 481 ctgtaaatgctgaataataacttatacttgaatgaagggttacaagacatgttgaactt 540
|||||
Db 4130 CTTGAATATGCTTAATATATCTTATGCAATTAAGGCTTACAGAACATGTTGAACATT 4189
|||||
Qy 541 ttttaacttatttgggaataagaatgtttgcacctccacattatgttctt 593
|||||
Db 4190 TTTTACTTTTATTTGGGAATTAAGGAATGTTTGCACCTCCACATTTATTTGCTT 4242
|||||

RESULT 2
AX119931
LOCUS AX119931 12793 bp DNA linear PAT 11-MAY-2001
DEFINITION Sequence 1 from Patent WO0129266.
ACCESSION AX119931
VERSION AX119931.1 GI:14036678
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE
1 (bases 1 to 12793)
AUTHORS Hudson,T.J., Engert,J. and Richter,A.
TITLE Identification of asrac mutations and methods of use therefor
JOURNAL Patent: WO 0129266-A 1 26-APR-2001;
MCGILL UNIVERSITY (CA) ; Hopital Sainte-Justine (CA)
Location/Qualifiers

FEATURES
source 1. 12793
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 4163 a 2256 c 2487 g 3887 t
ORIGIN

Query Match 100.0%; Score 593; DB 6; Length 12793;
Best Local Similarity 100.0%; Pred. No. 3.6e-92;
Matches 593; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 acatctatgttaccaggtccctgttgaagaagatgaacggaacaaactaaatgtt 60
|||||
Db 12201 ACATCTTAGTTTACAGGCTTCTGTTGATGAAGATAGCAAGCAAACTAAATGTT 12260
|||||
Qy 61 ggcagttcttattaccagttgttgaatgttcttgcgaacgtctgcgaagaacatt 120
|||||
Db 12261 GGCAGTTCTTATTACAGGTTGATGATGTTCTGCAAACTCTGCGCAAGCAACATT 12320
|||||
Qy 121 tattaactgttagaacaactgtcttattgttgtgtgttacaatcttccacaatgttata 180
|||||
Db 12321 TATTAACTGTTAGAACACTGCTTATGTTGTGTGTACATATTTCACAAATGTTATA 12380
|||||
Qy 181 attatatagttgttgaacagatgcaactcttctgttcttcaaaagtctgcagtttaa 240
|||||
Db 12381 ATTATATAGTGTGTGAACAGATGCAATCTTTGTTGTCTGAAGGCTCTGCAGTTAA 12440
|||||
Qy 241 aaaaaaaacaaactcttcttcaatagtgaatgttgcagttcttcaacttaaaac 300
|||||
Db 12441 AAAAAAAACAACTTTCTTCAATATGCGATGTAGGGGATTTTAACTTAAAAAC 12500
|||||
Qy 301 atcaaaaatgttaaaatcaatctgttatactagtagttataactacagctataattc 360
|||||
Db 12501 ATCAAAAATGTAAATCACTTTGTGTATGTAGTAGTTTAAATATCGGCTTATATTTC 12560
|||||
Qy 361 cccatgaatgacagactgacattaatcatgttctgcacatgctcttcaactt 420
|||||
Db 12561 CCAATGAATGATGAGAACGTGCAATTAATCAATGTTGTGTGTGCGCAAGCTTCTTTACTTT 12620
|||||
Qy 421 aacatattcttcttgaagaatgaaaggtaagatgaatgaattatataaagtactg 480
|||||

THIS PAGE BLANK (USPTO)

Wed May 22 09:23:25 2002

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model
May 22, 2002, 05:31:11 : Search time 3619.39 Seconds
(without alignments)
5224.426 Million cell updates/sec

Run on:
US-09-693-205-7_COPY_5300_6700
101
1 gtagcagtaaacatagagc.....aaagcacttaagaagt 1401
Sequence: 53rd
IDENTITY: NUC
Gapop 10.0, Gapept 1.0

Scoring table: 13736207 seqs, 6748477542 residues 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database:

1: em_estba:*
2: em_esthum:*
3: em_estlin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_hiv:*
15: em_gss_pla:*
16: em_gss_vit:*

Pred. NO. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Score	Match	Length	DB	ID	Description
1	752.2	53.7	1083	10	BM476887	BM476887 AGENCOURT
2	673.4	48.1	795	10	R17106	R17106 EST20108 C1
3	458.4	32.7	1010	10	BM472954	BM472954 AGENCOURT
4	345.2	24.6	376	9	AL596565	AL596565 DKFZP451L
5	226	16.1	238	9	AA897178	AA897178 am09608.s
6	203.4	14.5	209	9	AA776169	AA776169 ae08002.s
7	144.4	10.2	148	9	AA776670	AA776670 ae08002.s
8	114.8	8.2	505	9	AA217954	AA217954 SP_0148.B
9	107.8	7.7	434	9	AA776169	AA776169 SP_0148.B
10	87.6	6.3	1067	12	CNS02W80	CNS02W80 xw61a11.x
11	87.2	6.2	586	10	BF650306	BF650306 NF087011E
12	85	6.1	918	12	A2136287	A2136287 SP_0169.B
13	80	5.7	807	12	A2185191	A2185191 SP_0104.B
14	80	5.7	431	9	A2182120	A2182120 SP_0108.A
15	79.6	5.7	331	9	AT685137	AT685137 wa75a03.x
16	67.2	4.6	821	12	BH601255	BH601255 BCGW28TR

us-09-693-205-7_copy_5300_6700.std.rst

18	61.6	4.4	643	9	BB485245	BB485245
19	58.4	4.2	284	9	BB254184	BB254184
20	58.2	4.2	445	10	BF646344	BF646344
21	57.8	4.1	616	12	A2963183	A2963183
22	55.2	3.9	234	9	BB254737	BB254737
23	51.4	3.7	721	12	AA0254586	AA0254586
24	48.4	3.5	1069	12	CNS00396	CNS00396
25	47.8	3.4	936	12	CNS0276Y	CNS0276Y
26	47.2	3.3	959	12	CNS040NC	CNS040NC
27	46	3.2	1101	12	CNS0006J	CNS0006J
28	44.8	3.2	555	12	AQ310013	AQ310013
29	44.4	3.2	560	12	CNS040H	CNS040H
30	44.4	3.1	937	12	AM517860	AM517860
31	43.8	3.0	480	9	AM294501	AM294501
32	42.6	3.0	588	10	BM274569	BM274569
33	42.6	3.0	609	12	BE440138	BE440138
34	42.6	3.0	670	12	AC0308275	AC0308275
35	42.4	3.0	987	12	CNS0014PQ	CNS0014PQ
36	42.2	3.0	979	12	CNS0009H	CNS0009H
37	42.2	3.0	536	12	A0757450	A0757450
38	42	3.0	782	9	AL552607	AL552607
39	41.8	3.0	466	12	BH479411	BH479411
40	41.6	3.0	524	12	BH600863	BH600863
41	41.6	3.0	547	12	BH502803	BH502803
42	41.6	3.0	611	12	BH436907	BH436907
43	41.6	3.0	616	12	BH463467	BH463467
44	41.6	3.0				
45	41.6	3.0				

ALIGNMENTS

RESULT 1
LOCUS BM476887 1083 bp mRNA
DEFINITION AGENCOURT_6481789 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:5555441
5', mRNA sequence.
ACCESSION BM476887
VERSION BM476887.1 GI:18525929
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE NIH-MGC Human Genome Project (HGP)
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgap@nci.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
plate: LLAM12275 row: 1 column: 18
High quality sequence stop: 656.
Location/Qualifiers
1..1083
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5555441"
/clone="NIH_MGC_71"
/clone="lib=telomostroma"
/tissue="type=telomostroma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: uterus; Vector: pCMV-Sport6; Site: 1; NotI
Site: 2; SalI; Cloned unidirectionally. Primer: Oligo dT
Average insert size 2.1 kb.
334 a 219 c 212 g 318 t

FEATURES

source
BASE COUNT
ORIGIN

TCAACAATAACTATCTATGCTACTGAGGACCTGTGNAGGAATCTTTCCTGGTTGGT

us-09-693-205-7_copy-5300_6700.std.rst

```

CDNA Library Arrayed by: The Bioscience Corporation " can be
DNA Sequencing by: Agencourt Distribution Information at:
Clone distribution: MGC clone distribution information;
found through the I.M.A.G.E. Consortium/MLN at:
http://image.llnl.gov
Plate: LLM12323 row: h column: 13
High quality sequence stop: 738.
Location/Qualifiers
1. 1010
/organism="Homo sapiens"
/db xref="taxon:9606"
/clone="IMAGE:5574588"
/clone_id="NH_MGC_88"
/lisue_type="duodenal adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: small intestine; Vector: pCMV-Sport6;
site_1: NotI; Site_2: SalI; Cloned unidirectionally
oligo-dr primed. Average insert size 1.767 kb. Library
enriched for full-length clones and constructed by Life
Technologies . Note: this is a NH_MGC Library."
BASE COUNT      293 a      209 c      191 g      317 t
ORIGIN
Query Match          32.7%; Score 458.4; DB 10; Length 1010;
Best Local Similarity 99.4%; Pred. No. 3e-106; Indels 2; Gaps 2
Matches 481; Conservative 0; Mismatches 1;
QY    920   ccgttcacaacaaataacctactctcttgatgaactagaagcctctgaaaggaaattcttactactgcgt 979
       |||||||
Db     20    CAGTTCAACAATAAATCCTATTACTTACNGCATGATCGAGACTGTGAAGAAATCTTACTACGCT 79
OY    980   ggctcaatttgtaatatagatcaggcctttccaaatagtaggaagaagtacttaaagtgcatat 1039
       |||||||
Db     80    GGCTTAATTTGGTAATAGATCGCGCTTTTCAGAATATGAGAAAGATCTTAAAGTGTCATAT 139
OY    1040  cagctcacaaagaacaaagatattactctttcccacggcgtagtgagtagcgtgcgtcattra 1099
       |||||||
Db     140  CAGCTCACAAAGAACCAAGATATTTACTCTTTTCCACAGTGCTGGAGTAGACTGCCCTGCATTA 199

```

	RESULT	4	
AL596565	Al596565		Homo sapiens EST 14-AUG-2001
LOCUS DKEFP45IL010.r1_451 (synonym: hlcc1) spinal cord Homo sapiens cDNA	DKEFP45IL010.5' mRNA sequence.	376 bp	
DEFINITION	Clone DKEFP45IL010.5'		
ACCESSION	AL596565		
VERSION	AL596565.1 GI:15154261		
KEYWORDS	EST.		
SOURCE	human.		Euteleostomi;
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eutelestomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (bases 1 to 376)		Olsenwelder, B., Obermaier, B., Mwes, H.W., Well, B. and Wiemann, S.
TITLE	EST (Olsenwelder, B., Obermaier, B., Mwes, H.W., Well, B. and Wiemann, S.) unpublished (2001) Contact: Olsenwelder B MPS Am Kioferspitz 18a D-82152 Martinsried, Germany		
JOURNAL COMMENT	This is the 5' sequence of the clone insert. This is from S. Wiemann, Molecular Genome Analysis, Heidelberg.de: Research Center (DKFZ), Email s.wiemann@dkfz-heidelberg.de sequenced by Medigenomix (Martinsried/Germany). Within the German genome consortium of the German Genome Project. No 51 sequence available. This clone (DKEFP45IL010) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clonerzp.dg.de		
FEATURES	Location/Qualifiers		
source	1..376 /organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="DKEFP45IL010"		
	/clone_lib="451 (synonym: hlcc1)"		
	"spinal cord"		
	/tissue_type="human spinal cord"		
	/dev_stage="adult"		
	/lab_host="DHIOB"		
	/note="Vector: pSport1; Site_1: NotI; Site_2: SalI"		
BASE COUNT	127 A		
ORIGIN	63 C		
	107 T		
	79 G		
Query Match	24.6%	Score 345.2:	DB 9% Length 376:
Best Local Similarity	98.4%	Pred. NO. 1.ee-77:	Indels 3; Gaps
Matches 370;	Conservative	0; Mismatches	3;

	BEST LOCAL SIMILARITY	99.68%	Pred. No. 4,Se-47;	
	MATCHES 237; CONSERVATIVE	0;	MISMATCHES 0;	INDELS 1; GAPS 1
	QUERY 1156	actggcgctgccatttcacgttgatgtagtgcgaacctttgcactgatgataccagaagaaacctg	1215	
	DB 238	actggcgctgccatttcacgttgatgtagtgcgaacctttgcactgatgataccagaagaaacctg	1215	
	QY 1216	tggcgtatgataatgaggagtgtgcttcgcgaagtgatgtaatacaagctttaatgaagaca	179	
	DB 178	tggcgtatgataatgaggagtgtgcttcgcgaagtgatgtaatacaagctttaatgaagaca	179	
	QY 1276	tttaaaggctccctgcatatg- ⁴ tgaattgcttaatacaagctttaaaaaaacggcattccctggt	1334	
	DB 118	tttaaaggctccctgcatatg- ⁴ tgaattgcttaatacaagctttaaaaaaacggcattccctggt	1334	
	QY 1335	tctgctccaacatatacagtggttacagaacacccttatcatgcttgctlaagaagaacct	1392	
	DB 58	tctgctccaacatatacagtggttacagaacacccttatcatgcttgctlaagaagaacct	1392	
	RESULT 6	AAT76169		
	LOCUS	aes0c02.s1	Stratogene schizo brain S11 Homo sapiens cDNA clone	EST 05-FEB-1998
	DEFINITION	AAT76169		
	ACCESSION	AA776169		
	VERSION	AA776169.1	GI:2835503	
	KEYWORDS	EST.		
	SOURCE	Homo sapiens		
	ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
	REFERENCE	1 (bases 1 to 209)		
	AUTHORS	Hillier,L., Allen,M., Bowles,L., Dubouque,T., Gaisel,G., Jost,S., White,Y., Wille,T., Waterston,R., Steptoe,B., Tan,F., Theising,B., WashU-NCI human EST Project		
	TITLE	Unpublished (1997)		
	JOURNAL	Contact: Wilson RK		
	COMMENT	Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@wustl.edu This clone is available royalty-free through LNLN; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Seq primer: -40m1 fwd. RT from Amersham High quality sequence stop: 206. Location/Qualifiers 1..209		
	FEATURES			
	SOURCE			

```

Forest Park University School of Medicine
314 286 1800
314 286 1810
: est@watson.wustl.edu
Clone is available
Consortium (info@clone.royalty-free through LINT ; contact the
clone reversed clone: info@clone.111.gov) for further information.
timer: -40m13 fwd. 87 from Amersham
Nullity sequence stop: 206.
Location/Qualifiers
1. 209
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:970466"
/clone_11b="Stratagene schizo brain s11"
/sex="male"
/tissue_type="schizophrenic brain s-11 frontal lobe"
/dev_stage="34 years old"
/lab_host="SOLR (kanamycin resistant)"
/notes="Vector: Bluescript SK-, Site_1: EcoRI. Library
constructed from S-11 frontal lobe, male, 34 years old,
50% caucasian, 50% Aloutian. Schizophrenic suicide.
Random primed into EcoRI site of ZAP II Vector.
excised. Custom library. Avg insert length 1.4kb.
Material obtained by Johnson N., Torrey, E.F., Volken R.,
and the Stanley Neuropathology Consortium - Analysis of
RNAs from the Brains of Individuals with Psychiatric

```

FEATURES	source	location/Qualifiers
	1..499	/organism="Strongyloides purpuratus" /dp.xref="taxon:7668" /clone="plate-148 Col-12 Row-N" /clone_1lp="Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library" /note="Organ: sperm; Vector: BAC3.6; BAC Clones in E-Coli DH10B"
BASE COUNT	140 a	96 c 116 g 146 t 1 others
ORIGIN		
Query Match	9.6%	Score 134.6; DB 12; Length 499;

Best Local Similarity 61.3%; Pred. No. 1.1e-23;
Matches 249; Conservative 0; Mismatches 155; Indels 2; Gaps 2;

QY 134 ttaatcatatccttcttgaagaagaatgttgaagaagcttcaaaagctgatgatg 193
1 tttatctatcaccgtcttgatgaaacatcttanaagactcttccagatgcatg 60
QY 194 caaaggcgaagaagaatcgttctgttctgacacagacacacagttgataaat 253
Db CTGGGCACTGAATCCATCTTGTCTAGCATATGCGCTTACATAGCGAAATTAACCTTC 120
QY 254 ttgatgataagctggcccatctgcaaggccagacttctgtgtacacacacagccat 313
Db 121 TAAGAAATGACATGAAGCTTTCAGGAGCTCTCTTGTCTTCAACACAGACTACT 180
QY 314 ttcaagaagatgattgttagaagattcagactcttgaaagacagaaaggaggaatc 373
Db 181 TTACAGATGACGATATTCGTGGGATACAGATCTTGACAAAGTACGTAAGCTGATCACA 240
QY 374 ctataaa-actggaagatgataagataagataatctgtgacatcatatcagactgc 432
Db 241 TATGAAAGATGAAGCCGTTATGCGGTTCATGCGGTATCAACGCTACACGAGATTCG 300
QY 433 ccaatcttcttcttctgcaatgacatccctgtgattttagatcccatcagatatgca 432
Db 301 CCAAGTTTCATCTTCATATGCTGACCCGATTAATGCGATGTTGACCTTC-TGTTGAATATAT 359
QY 493 ccaggggccacatccatagtcacggagcagatctttagaagatttg 538
Db 360 CCGGAGACCAACCCGATTAAGCCGCGCTTCACATATTAAGATATGG 405

RESULT 9
AM295072 505 bp mRNA linear EST 16-JAN-2000
LOCUS
DEFINITION UI-H-BW0-ait-c-04-0-UI.s1 NCI-CGAP_Sub6 Homo sapiens cDNA clone
IMAGE:2730247 3', mRNA sequence.
ACCESSION AM295072
VERSION
KEYWORDS
SOURCE EST.
ORGANISM human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 505)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
CONTACT: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Oligo-dT track not found. Not 1 site shown in beginning of sequence
is likely internal to the message. cDNA library preparation: M.B.
Soares Lab Clone distribution: NCI-CGAP clone distribution
Information can be found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/dbp/image/image.html
Seq primer: M13 Forward
POLYA-No.

FEATURES

Location/Qualifiers
1..505
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2730247"
/clone.lib="NCI CGAP Sub6"
/lab_host="PH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not 1; Site_2: Eco RI; NCI-CGAP_Sub6
is a subtracted library derived from BW, which consists of
a mixture of four normalized libraries: NCI CGAP_Bn50,
NCI CGAP_Ln13, NCI CGAP_Ov18, GBC1. The NCI-CGAP_Sub6
library had 7 million recombinants. A single-stranded DNA
preparation of BW was used as a tracer in a subtractive
hybridization with a driver comprising: the IMAGE pool

(NCI CGAP_Kid3 pool 1 LIAM 3334-3337, 3682-3683,
3798-3803 (IMAGE Clonoids 1322376-1322911,
1456008-1456775, 1500552-1502855); NCI CGAP_Kid5 pool 1
LIAM 3338-3342, 3722-3725, 3776-3778 (IMAGE Clonoids
1323912-1325831, 1473368-1472903, 1493104-1493255);
NCI CGAP_Ln5 pool 1 LIAM 3575-3582, 3851-3854 (IMAGE
Clonoids 1414920-1417991, 1520904-1522439); NCI CGAP_Gc4
pool 1 LIAM 3164-3167, 3716-3720, 3733-3735 (IMAGE
Clonoids 1257096-1258631, 1469064-1470983, 1475592-1476743
); NCI CGAP_Pr22 pool 1 LIAM 2457-2459, 2758-2759,
3062-3068 (IMAGE Clonoids 985608-986759, 1101197-1101959,
1217928-1220615); NCI CGAP_Co10 pool 1 LIAM 2644-2653,
2871-2872 (IMAGE Clonoids 1057416-1061255, 1144584-1145351
). (50% of the driver population), plus a pool of 3,840
arrayed clones from NCI CGAP_Sub1 (IMAGE Clonoids
2708616-2710535) and NCI CGAP_Sub2 (IMAGE Clonoids
2710536-2712455) (20% of the driver population), plus a
pool of 11,136 clones from NCI CGAP_Sub3 (IMAGE Clonoids
2712456-2723591) (30% of the driver population).
Subtraction was performed as previously described [Bonaldo
& Lennon & Soares (1996): Normalization and Subtraction:
Two Approaches To Facilitate Gene Discovery. Genome
Research 6, 791-806
TAG_Lib=NCI CGAP_Pr22
TAG_Tissue=Prostate
TAG_SEQ=AAATG"

BASE COUNT 159 a 93 c 117 g 136 t
ORIGIN

Query Match 8.2%; Score 114.8; DB 9; Length 505;
Best Local Similarity 57.5%; Pred. No. 1.3e-18;
Matches 206; Conservative 0; Mismatches 152; Indels 0; Gaps 0;

QY 122 ttaagagcatccttaatgcatatccttctgaaagaagaatgttgaagaagcttcttcaaa 181
Db 138 TCAAGGACATCTTTGACAGAGATCCAGAAAGGAGACAGATTCCTTAAGATTAATTCACA 197
QY 182 atgctgatgatgaaagagcagagaatcgttctgttctgacacagacatccag 241
Db 198 ATGCAGAGATGCTGTGGGGCCACAGAAAGTTAATTTTATATGATGAACCTATACGGA 257
QY 242 ttgatagaatatttgatgaatgaatggcccatgcaaggccagacattgtgttaca 301
Db 258 CACAGACTCTTTGGTCAAAAAGATATGCGCCATATGAGGGCCAGCTCTATGTGTACA 317
QY 302 acacaagccattacaagaatgattgtagaagaattcagaatcttgaaagaagcaga 361
Db 318 ACAAGCGGCTTTCACCCGAGAGACTGGCAGCATTCAGAAATAGCAAGAGACGAGA 377
QY 362 aagaggaagaatccttataaagcagagatggaatagatcaatctgtgtatcata 421
Db 378 AAAAGATGATCTCTCGAAGGTGCGAAGATTTGGATTTGGTTATTTCTGTCTATCA 437
QY 422 tcacagactgcccacatcttattcttgcgaatgacatcctgtgtattttgacatca 479
Db 438 TAACAGATGTTCTTGTATCTTAGTGGGACCAAAATCCGAGATCAAAATCCATCA 495

RESULT 10
AM470182 434 bp mRNA linear EST 24-FEB-2000
LOCUS
DEFINITION xw61a11.x1 NCI CGAP_Pan1 Homo sapiens cDNA clone IMAGE:2832476 3',
mRNA sequence.
ACCESSION AM470182
VERSION
KEYWORDS
SOURCE EST.
ORGANISM human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 434)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL	Unpublished (1997)
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nih.gov Life Technologies catalog #: 11548-013 DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNLW at: www-bio.lnl.gov/dbfp/image/image.html Seq primer: -40up from Gidco High quality sequence stop: 412.
FEATURES	Location/Qualifiers
source	1..434
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/clone="IMAGE:2832476"
	/clone_lib="NCI CGAP, Pan1"
	/tissue_type="adenocarcinoma"
	/lab_host="DH10B"
	/note="Organ: pancreas; Vector: pCMV-SPORT6; site_1: SalI; site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.72 kb. Life Technologies catalog #: 11548-013"
BASE COUNT	141 a 76 c 101 g 116 t
ORIGIN	
Query Match	7.7%; Score 107.8; DB 9; Length 434;
Best Local Similarity	58.6%; Pred. No. 7.5e-17;
Matches 187;	Conservative 0; Mismatches 132; Indels 0; Gaps 0;
OY	122 ttaagagcatcccttaatgcatatccctcttgaaaggaagtgtgaagaagctcttcaaa 181
DB	115 TCACAGACATTTTGGAGAGATATCCAGAGAGGAGCAGACATCTTAAAGGAATTAATTCAGA 174
OY	182 atgctgcatgagtcgaagagcgacagaacatctgtttgttgttgatcccttgacagcatcaag 241
DB	175 ATGCGAAGATGCTCTGGCGACGAGAAAGTTAAATTTTATATCATGATAAATCGGAA 234
OY	242 ttgatagatatattgatgataagtcggcccccttgcaaggcgccagcacttgggtgtaca 301
DB	235 CAGACACTTTTGGTCAAAAGATATGGCGCCATATCAGGGGCCAGCTCTATGTGTACA 294
OY	302 acaacagccattacagaaagatgtagtgcgaaggaattcagaatcttgaaaaagcgaca 361
DB	295 ACAACGCGGTTTTCACCCACAGGAGACTGGACGCGATTCAAGAAATAGCAACAGCAGCA 354
OY	362 aagaggaatccctataaacctgacagcatggaatggattcaatctgtgtatcata 421
DB	355 AAAAGGATGATCCTCTGGAAGGTCGGAAGATTTGGATTGGTTAATTCGTCTATCATTA 414
OY	422 tcacagactgcacactctt 440
DB	415 TTAACAGATGTTCTGTAT 433
RESULT 11	
CNS02W80/c	1067 bp DNA linear GSS 15-MAR-2000
LOCUS	
DEFINITION	Tetradon nigriviridis genome survey sequence PUC-Or1 end of clone 176G09 of library G from Tetradon nigriviridis, genomic survey sequence.
ACCESSION	AL216729.1 GI:7875548
VERSION	AL216729
KEYWORDS	GSS: genome survey sequence.
SOURCE	Tetradon nigriviridis.
ORGANISM	Tetradon nigriviridis
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.
REFERENCE	1 (bases 1 to 1067)

AUTHORS		Roeest-Crollius,H., Jalllon,O., Dasliya,C., Fizemes,C., Fisher,C., Weissenbach,J., Billault,A., Queller,F., Saurin,W., Bernot,A. and Bouneau,L.,
TITLE		Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis
JOURNAL REFERENCE		Unpublished
AUTHORS		2 (bases 1 to 1067)
TITLE		Roeest-Crollius,H., Jalllon,O., Dasliya,C., Bouneau,L., Fisher,C., Bernot,A., Fizemes,C., Winker,P., Brotier,P., Queller,F., Saurin,W. and Weissenbach,J.
JOURNAL REFERENCE		Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence
AUTHORS		Unpublished
TITLE		Genoscope. Submitted (12-Apr-2000) to the EMBL/Genbank/DBJ databases This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetraodon.Location/Qualifiers
FEATURES		source 1..1067 /organism="Tetraodon nigroviridis" /db_xref="taxon:99883" /clone="I76G09" /clone_lib="G" /note="Genoscope sequence ID : COAG176AD05SP1-end : PUC Ori"
BASE COUNT		250 a 249 c 239 g 328 t 1 others
ORIGIN		
Query Match		6.3% Score 87.6; DB 12; Length 1067;
Best Local Similarity		54.3%; Pred. No.1,3e-11;
Matches 202:		Conservative 0; Mismatches 164; Indels 6; Gaps 1;
Oy	87	agaatttgaggcagaaaataatgacgcagaaattaagagcatcttaatgatcatcc 146
Db	626	ACAATGCGACAAATCGAACCCGATTACCAGAGAGATAAAAAAATATTCTGAACGATACGA 567
Oy	147	tctgtgaaaaaggaagtgtgaagaagactttccaanaatgcttgtatgttcgaaggcgacga 206
Db	566	TCAGGAGGGGTGCATCTTTAAAGAGCTCATCAAAAACGAGAAGATGCTGGTCGCAGACA 507
Oy	207	aatctgttctgtgtgttatccttagaacagcatcaagttgtagaaa-----tatctgatga 260
Db	506	ATGCTTGTTCTCTGTGGTAGATTTCAGAACGCCAACAATTAATTCCTGAGAGACTATTTGACCC 447
Oy	261	taagtgggcccatitgcaaaggcgaccatttgytltgtaacaacacagccalttacaga 320
Db	446	CGACATGAGACTCCTGTGACAGGGCCCTTGCTGTGGGGCTTAAATATGACAGCTCAAGA 387
Oy	321	agatgatgttagaagaaatcagaatccttggaaaaggcagcaaaaggaggaatatccttata 380
Db	386	TCAGCAGCTGCAGAACAACTTCTCCTGGTGGCTCTGCTCTAAAGGAACAACCACTGGAGAA 327
Oy	381	aacttgacacgtatgtaataagattcaattctgtatcatatcacagactgccatcttt 440
Db	326	ATATCGAAAGTTTGCGCTTGATTCACACCTGTAATCATGTGACAGACATCCCTTCAT 267
Oy	441	tatttctgcaaa 452
Db	266	CCTCAGTGGAAA 255
RESULT 12		BF650306
LOCUS		BF650306 586 bp mRNA linear EST 20-DEC-2000
DEFINITION		NF087D11EC1F1092 Elicited cell culture Medicago truncatula cDNA clone NF087D11EC 5', mRNA sequence.
ACCESSION		BF650306
VERSION		BF650306.1 GI:11915436

KEYWORDS	EST.
SOURCE	barrel medic.
ORGANISM	Medicago truncatula
	Eumariotia, Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
	Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliales;
	Medicago.
REFERENCE	1 (bases 1 to 586)
AUTHORS	Torres-Jerez, I., Scott, A.D., Harris, A.R., Gonzales, R.A., Bell, C.J.,
	Flores, H.R., Inman, J.T., Meller, J.W. and May, G.D.
TITLE	Expressed Sequence Tags from the Samuel Roberts Noble Foundation -
	Center for Medicago Genomics Research
JOURNAL	Unpublished (2000)
COMMENT	Contact: Dixon RA
	Plant Biology Division
	The Samuel Roberts Noble Foundation
	2510 Sam Noble Parkway, Ardmore, OK 73402, USA
	Tel: 580 221 7302
	Fax: 580 221 7380
	Email: radixon@noble.org
	Insert length: 586 std Error: 0.00
	Plate: 087 row: D column: 11
	Seq primer: TCACACAGAAACACTATGAC.
FEATURES	Location/Qualifiers

FEATURES	SOURCE	LOCATION/QUALIFIERS
		1. 586
		/organism="Medicago truncatula"
		/db_xref="taxon:3880"
		/clone="NF087D11EC"
		/clone_lib="Elcited cell culture"
		/tissue_type="Cell cultures derived from root tissues"
		/dev_stage="Cell suspensions were subcultured every 14 days. Cells were induced six days after subculture"
		/note="Vector: lambda zap; Cells were induced with yeast cell wall extracts equivalent to 50ug/ml glucose in the final concentration. Samples were taken at 0.5', 1', 12 and 24 hours after induction. Equal amounts of RNA from each time point were pooled and used for mRNA isolation."
BASE COUNT		135 a 151 c 127 g 171 t 2 others
ORIGIN		

Query Match	6.2%	Score 87.2	DB 10	Length 586
Best Local Similarity	51.2%	Pred No. 1.4e-11		
Matches 228, Conservative	0	Mismatches 214	Indels 3	Gaps 1

QY	74	caacattcgacacgaatttgcgagaaagaaattgacgcgaattgaagcatcc	133
Db	76	CAATTTCTTCGAGATTTCGGCCAAACAGTCGACCTAATCTGTGCACTCCGGAATTC	135
QY	134	ctaaatcatatcctctctgaaagaagatgttgaagagctcttcaaaatgtcatgtatg	193
Db	136	TCTTAAATCTATCCGAAGGAACACCGTTCCTCAAGAGAGCTTATTCAGAACGCCGATGAG	195
QY	194	caaaaggcagcaagaatctggttttctgtttgatcctagacagcatccagttgatagaatc	253
Db	196	CCGGTGCTACCAACCGTTTCTCTCTCTCTGTGAGACGGCCGTCACACAGCGCCGTGATGCTGTAC	255
QY	254	ctgatgataaagtggcccccatcgacaggcgacagacttgtgtgtaacaacacagcaat	313
Db	256	TGTCCGATTTCCGTATTCGACAGTGGCAAGACCTTCGCCCTTATGATATGATGCTGTTT	315
QY	314	ttacagaagaatgattagagaatttcagaatcttgcgaaagcgacgaagaagggaatc	373
Db	316	TCTCTGAGAGAGATTGTTGTAAGTATTCGAAGATCGGGGAGTGTGTAAGATGGACAG	375
QY	374	cttataaaactggacagfatggaatagaattcaattctgtgtatcataatcacagactgcc	433
Db	376	CTTCCAAAACCGGTGATTCGGGGTTGGCTTCACCTCAGTGCATTTAAACAGATCTTC	435
QY	434	catctttatcttgcgaaatgacatcccgtytaatttttgatccctatgcagatgatgac	493
Db	436	CTTCAATTGCGTAGTGGCAAAATCTTGAGTAT--TGTTTGAACCTCAAGGTGTTTATCTTTC	492

Qy	494	caggcgccatccatctagtc	ccg	518
Db	493	CAAGGTTCTGCANCAATCCTG		517

RESULT 13
AZ136287

LOCUS	AZ136287	918 bp	DNA	linear	GSS 28- AUG-2000
DEFINITION	SP_0169_B2_B09-SP6E Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library Strongylocentrotus purpuratus genomic clone Plate=169 Col=18 Row=D, DNA sequence.				

ACCESSION	AZ136287
VERSION	AZ136287.1
KEYWORDS	GI:8288190
	GSS.

SOURCE ORGANISM	Strongylocentrotus purpuratus.
	Strongylocentrotus purpuratus

Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa; Echinoida; Echinoidea; Euechinoidea; Echinacea; Echinoida;

Strongylocentrotidae; Strongylocentrotus.
1 (bases 1 to 918)

AUTHORS
Cameron, R.A., Mahatiras, G., Raat, J.P., Martinez, P., Blondi, T.R., Swartzell, S., Wallace, J.C., Poustka, A.J., Livingston, B.T., Wray, C. & Etzel, C. A. (1996) *Journal of Interpersonal Violence* 11, 103-114.

TITLE A sea urchin genome project: Sequence scan and virtual map
G.A., Eltensohn, C.A., Lemach, H., Bitten, K.O., Davison, E.H. and Hood, L.

JOURNAL Proc. Natl. Acad. Sci. U. S. A. 97 (17), 9514-9518 (2000)

COMMENT	Contact: Cameron, RA, Davidson, EH, Hood, L
MEDLINE	20402566

Division of Biology 156-29
California Institute of Technology

Pasadena California 91125, USA
Tel: (626) 395-8421

Fax: (626) 793-3047
Email: acameron@caltech.edu

Plate: 169 row: D column: 18
Seq primer: SP6

Class: BAC ends
High quality sequence stop: 918

```

FEATURES
source
location/qualifiers
1. .918
/organism="Streptococcus"

```

```

/organism="Strongyloides purpuratus"
/db_xref="taxon:7668"
/clone="plate169 COL18 RowD"

```

urchin, sperm genomic BAC library"
/clone_lib="Strongylocentrotus purpuratus, purple sea
urchin, P146-105 COL-10 ROW-D"

acquiring sperm genomic DNA library
/note="Organ: sperm; Vector: BACe3.6; BAC Clones in E-Clon
DH10B"

BASE COUNT	262 a	207 c	214 g	235 t
ORIGIN				

Query Match	6.18;	Score 85;	DB 12;	Length 918;
Best Local Similarity	54.08;	Pred. No. 5.7e-11;		

Category	Count
Matches	222
Mismatches	180
Indels	9
Gaps	2

[illegible][illegible]

140 CATTATGACAAACCCTAATCCCTTTCAGTAABATCAATCAGAAAGAACGCATGA 100
+ | | | | | | | | | | | | | | | |
144 CCCCTCCAAAAGAATGTCGGAAGAGCTCTCTCCAATAATCATGATGACGAAGCGCAC 203
+ | | | | | | | | | | | | | | | |

304 aaaaatctatttatatttaactctacacacacacatccaatttaaf-----aaaaattatttga 357
DD CTAATAAGCACCCGCTATCGTAAAGTAAATGATACAGAAATCGCGAGATATGCAAGAGCTCA 139
140

200 GAGATTCGCTTTCATTTGACATGAGAGAAAGCTTTTAGAACCACACAGATTGTTGGA 259

258 tatataaattgcacccatttcacaaagccacactttatctgtatcacaaacacccacattac 312

Db 260 TCCGGGAATGAAATCTTGTCAAGGACCAACCCCTTGGGATATACAACGATGCCGTATTCAC 319

318 agaaagatgattttaagaggaatlcagaatccttgaaaaagcacgaaagaagaggaatcctta 377

Db 320 AGACCAAGACTTGGAGATATTTCTTCTTGGAGGGGCTACCAAGAAAGACGCTGA 379

Qy 378 taaactgcagatgataatgaatgaatcattcgtatcatatcacagactcccatc 437

Db 380 AAAGATGTTGGAGATTTGGACCGCTTCAACTCGGTACCGCATTAATGTTCCAA 439

Qy 438 ttattatcggcaatgacatccgtgtattttgatccatcgtacgata 488

Db 440 TTTGTTAGTAGGAGCTACATGAGT---TTTTCGACCCCTCATACACACA 487

RESULT 14

AZ185191 807 bp DNA linear GSS 30-AUG-2000

LOCUS SP_1004_B2.D05_T7A Strongylocentrotus purpuratus, purple sea urchin

DEFINITION 'sperm genomic BAC library Strongylocentrotus purpuratus genomic clone plate-1004 Col-10 Row-H, DNA sequence.

ACCESSION AZ185191

VERSION AZ185191.1 GI:8357566

KEYWORDS GSS.

SOURCE Strongylocentrotus purpuratus.

ORGANISM Strongylocentrotus purpuratus

Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa; Echinoidea; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae; Strongylocentrotus.

REFERENCE 1 (bases 1 to 807)

AUTHORS Cameron, R.A., Mahairas, G., Rast, J.P., Martinez, P., Biondi, T.R., Swartzell, S., Wallace, J.C., Poustka, A.J., Livingston, B.T., Wray, G.A., Etlensohn, C.A., Lehrach, H., Britten, R.J., Davidson, E.H. and Hood, L.

TITLE A sea urchin genome project: Sequence scan, virtual map, and additional resources

JOURNAL Proc. Natl. Acad. Sci. U. S. A. 97 (17), 9514-9518 (2000)

MEDLINE 20402566

COMMENT Contact: Cameron, RA, Davidson, EH, Hood, L

Division of Biology 156-29

California Institute of Technology

Pasadena California 91125, USA

Tel: (626) 395-8421

Fax: (626) 793-3047

Email: acameron@caltech.edu

Plate: 1004 row: H column: 10

Seq primer: T7

Class: BAC ends

High quality sequence stop: 807.

FEATURES

source

1. 807

Location/Qualifiers

/organism="Strongylocentrotus purpuratus"

/db_xref="taxon:7668"

/clone="Plate-1004 Col-10 Row-H"

/clone_lib="Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library"

/note="Organ: sperm; Vector: BAC3.6; BAC Clones in E-Coli DH10B"

BASE COUNT 245 a 182 c 186 g 192 t 2 others

ORIGIN

Query Match 6.0%; Score 84; DB 12; Length 807;

Best Local Similarity 54.2%; Pred. No. 1e-10;

Matches 218; Conservative 0; Mismatches 175; Indels 9; Gaps 2;

Qy 93 tgggcagaagaatgacagcagaatgaagcattccttaatcctctctga 152

Db 35 TGGGCACATGACCTTGAATGCGCTGTGGAATATCTCAAGAACATATGTGGA 94

Qy 153 aaaggaattgtgaagaagcttctcaaatgctgatagtcaagcgacagaatctg 212

Db 95 CACGCGCATGTTAGTAAATGATACAGAAATGACAGAGATCAGAGCTCAGGAAGTTCG 154

Qy 213 ttctgtttgacatcctgac-----agcatcaggttgatagaatatttgatagtg 266

Db 155 CTTTCTCATGTGACATGACAGAAAGACGAGAAACCAAAAGATTATTCATCCGGAA 214

Qy 267 gcccacattgcagaagccagcacttctgtgtacaacaacagcatttaagaagatga 326

Db 215 GAAGCTCTGTGAGGACACCGCCTTTGGGTATACAAACATCTGTATTCACAGACAGA 274

Qy 327 tgttagaggaattcgaatcttggaaaagcgacgaagaagggaatctcttaaacg 386

Db 275 CTTTAAAGATATTTCTCCGCTTTGGAGGGCTACCAAGAAAGACGCTGAAACATCGG 334

Qy 387 acagatggaatagatcaatcctcgtatcatatcacagactgccatcttatttc 446

Db 335 GAAGTTTGGACCGCTTCAACTCGGTCTACCGCATTAATGTTCCAAAGTTTGCTTAG 394

Qy 447 tggcaatgacatcctgtgtattttgatccatcgtacgata 488

Db 395 TAGGAGACTACATGAGT---TTTTCGACCCCTCATACACACA 433

RESULT 15

AZ182120 431 bp DNA linear GSS 30-AUG-2000

LOCUS SP_0188_A1.A09_SP6E Strongylocentrotus purpuratus, purple sea urchin,

DEFINITION 'sperm genomic BAC library Strongylocentrotus purpuratus genomic clone plate-188 Col-17 Row-A, DNA sequence.

ACCESSION AZ182120

VERSION AZ182120.1 GI:8354495

KEYWORDS GSS.

SOURCE Strongylocentrotus purpuratus.

ORGANISM Strongylocentrotus purpuratus

Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa; Echinoidea; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae; Strongylocentrotus.

REFERENCE 1 (bases 1 to 431)

AUTHORS Cameron, R.A., Mahairas, G., Rast, J.P., Martinez, P., Biondi, T.R., Swartzell, S., Wallace, J.C., Poustka, A.J., Livingston, B.T., Wray, G.A., Etlensohn, C.A., Lehrach, H., Britten, R.J., Davidson, E.H. and Hood, L.

TITLE A sea urchin genome project: Sequence scan, virtual map, and additional resources

JOURNAL Proc. Natl. Acad. Sci. U. S. A. 97 (17), 9514-9518 (2000)

MEDLINE 20402566

COMMENT Contact: Cameron, RA, Davidson, EH, Hood, L

Division of Biology 156-29

California Institute of Technology

Pasadena California 91125, USA

Tel: (626) 395-8421

Fax: (626) 793-3047

Email: acameron@caltech.edu

Plate: 188 row: A column: 17

Seq primer: SP6

Class: BAC ends

High quality sequence stop: 431.

FEATURES

source

1. 431

Location/Qualifiers

/organism="Strongylocentrotus purpuratus"

/db_xref="taxon:7668"

/clone="Plate-188 Col-17 Row-A"

/clone_lib="Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library"

/note="Organ: sperm; Vector: BAC3.6; BAC Clones in E-Coli DH10B"

BASE COUNT 126 a 93 c 105 g 105 t 2 others

ORIGIN

Query Match 5.7%; Score 80; DB 12; Length 431;

Best Local Similarity 51.8%; Pred. No. 9.2e-10;

Matches 207; Conservative 0; Mismatches 187; Indels 6; Gaps 1;

Qy 84 cagcaaatgtggcagaagaatgacagcagaatgaagcattccttaatcctata 143

Db 9 CACACCACTGGCACAATGAACCTTGACTATGCGCTTGGAGATATTTCCACAGACAA 68

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 22, 2002, 07:30:25 ; Search time 2968.03 Seconds

(without alignments)
6370.971 Million cell updates/sec

Title: US-09-693-205-7_COPY_5300_6700

Perfect score: 1401

Sequence: 1 gtagcagcagtaaacatagagc.....aaagacacttaagaagt 1401

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 13736207 seqs, 674847542 residues

Word size : 0

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estcin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gD_est1:*
10: gD_est2:*
11: gD_htc:*
12: gD_gss:*
13: em_gss_hum:*
14: em_gss_iny:*
15: em_gss_pln:*
16: em_gss_vtl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	656	46.8	1083	10	BM476887 AGENCOURT
2	312	22.3	376	9	AL596565 DKE2P451L
3	205	14.6	1010	10	BM472954 AGENCOURT
4	201	14.3	209	9	AA776169
5	169	12.1	795	10	R17106
6	140	10.0	148	9	AA776670
7	140	10.0	238	9	AA897178
8	7	1.6	665	12	BH157201
9	21	1.5	247	9	AV214110
10	21	1.5	605	12	BH034007
11	21	1.5	517	10	BG788610
12	21	1.5	758	12	BH034007
13	20	1.4	227	10	BG624213
14	20	1.4	308	9	A1705955
15	20	1.4	320	9	A1579360
16	20	1.4	331	10	B1295731
17	20	1.4	355	9	AM435291

18	20	1.4	374	9	AM520724
19	20	1.4	383	9	BE101372
20	20	1.4	384	9	AM433545
21	20	1.4	418	12	AQ153679
22	20	1.4	461	9	A1043792
23	20	1.4	464	9	A1704578
24	20	1.4	477	9	AA988882
25	20	1.4	477	9	A1710607
26	20	1.4	517	10	BM287243
27	20	1.4	522	10	BE912359
28	20	1.4	534	10	B1612760
29	20	1.4	535	10	BM257525
30	20	1.4	541	10	BE668948
31	20	1.4	543	9	BE103356
32	20	1.4	546	12	AQ534831
33	20	1.4	563	10	B1290791
34	20	1.4	568	9	AL638993
35	20	1.4	571	12	AQ485516
36	20	1.4	616	12	AZ963183
37	20	1.4	621	12	AQ446973
38	20	1.4	652	10	BJ093045
39	20	1.4	693	9	BB365802
40	20	1.4	695	9	AL636882
41	20	1.4	701	9	AL631304
42	20	1.4	705	12	AQ255211
43	20	1.4	930	10	BF163742
44	20	1.4	931	12	AZ204179
45	19	1.4	106	12	AQ283816

ALIGNMENTS

RESULT 1
BM476887 1083 bp mRNA linear EST 05-FEB-2002
LOCUS AGENCOURT_6481789 NIH_MGC_71 Homo sapiens CDNA IMAGE:5555441
DEFINITION 5', mRNA sequence.

ACCESSION BM476887
VERSION BM476887.1 GI:18525929
KEYWORDS EST.

SOURCE

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE NIH-MGC http://mgc.nci.nih.gov/.

AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)

TITLE Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgabbs-remail.nih.gov

Tissue Procurement: ATCC

CDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC Clone Distribution Information can be

found through the I.M.A.G.E. Consortium/ILNL at:

http://image.llnl.gov

Plate: LAM12275 row: j column: 18

High quality sequence stop: 656.

Location/Qualifiers

1. 1083

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:5555441"

/clone_lib="NIH_MGC_71"

/tissue_type="leiomysarcoma"

/lab_host="DH10B (phage-resistant)"

/note="Organ: uterus; Vector: pCMV-SPORE6; Site_1: NotI;

Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.

Average insert size 2.1 kb.

334 a 219 c 212 g 318 t

BASE COUNT

ORIGIN

Query Match 46.8%; Score 656; DB 10; Length 1083;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 656; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 640 gcgaagaatgcaaaagtttcggaatttcgtctgttcacagatcagacagaatgtccag 699
 |||||||
 DB 13 GCAGAAATGCGAAAGTTTGGAAATTCGCTCTCCAGCATCAGACAGATGCTCAG 72
 QY 700 aatctttgcaaacctgcctcagatcgagcagaacttcattgttcttaacacatg 759
 |||||||
 DB 73 AATCTTTGGCAACACGCGCTCAGATGGGCGACAACCTCTAATGTTTCTTAACACATG 132
 QY 760 gaaaaattctattctgttgaataatagataagatcctgagctcctaattgtgtctatca 819
 |||||||
 DB 133 GAAAAATTTCTATTGTAATAGATAGAGTACTGAGCTCTAATATGCTCTATTCA 192
 QY 820 gtaaaaggcaaaatcacagatcgagacagatgtgaaaaggaaattctatgtctgtta 879
 |||||||
 DB 193 GTAAGGGCGCAAAATCACAGATGAGACAGATTGAAAAGAAATTCATGACATCTGTA 252
 QY 880 attgatgttctactaaagaagcagctcaagaacataccagttcaacaataacctat 939
 |||||||
 DB 253 ATTGATAGTGTACTATAAAGAGGACAGCTCAAGACATACAGTTCAACAAATACCTAT 312
 QY 940 actatgatactgagactcctgaagaatcttactacgtgtgtctaattgttaataatca 999
 |||||||
 DB 313 ACTATGATAGTACTGAGACTCTGAAGAAATCTTACTACGTGCTAATTTGTAATAGATCA 372
 QY 1000 ggccttcaagatgaggaagatatactaaagtgtcatatcagttcaagaaccaagat 1059
 |||||||
 DB 373 GGCCTTCAAGTATGAGAAAGATCTAAAGTGTCAATCAGCTCAACAAACCAAGAT 432
 QY 1060 attactcttccacacgtgtgtgagatgctgcctgcattactcacaactaaataaaacc 1119
 |||||||
 DB 433 ATTACTCTTTCCACGCTGGTGGAGTAGCTGCTGCATTATCTCAACTATAAANACC 492
 QY 1120 catagggcctctgtcttctgtcctctctcttgagacgtggcgtgcattcatgtgaat 1179
 |||||||
 DB 493 CATAGGGCCTCTCTCTTTTTCCTCTCTTCTTGGAGACTGGCGTCCATTTTCATGTGAAT 552
 QY 1180 ggcacatttgacactgattcagccagaagaactctgtggcgtgtatgtaagggttgc 1239
 |||||||
 DB 553 GGCACACTTGGACTGATTCAGCCAGAAAGAACTGTGGCTGATGATTAAGAGGTTGTG 612
 QY 1240 gtccgaagtgactggaataacagtttaatagacacattatagtcctgtcatcgt 1295
 |||||||
 DB 613 GTTCGAGTGACTGGAATACAGTTTAATGACACATTAAAGCTCCTGCATATGT 668

RESULT 2
 AL596565 376 bp mRNA linear EST 14-AUG-2001
 LOCUS DKEFZ451L0110.F1.451 (synonym: hlccl) spinal cord Homo sapiens cDNA
 DEFINITION clone DKEFZ451L0110.5, mRNA sequence.
 ACCESSION AL596565
 VERSION AL596565.1 GI:15154261
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 376)
 AUTHORS Ottenwaelder, B., Obermaier, B., Mewes, W., Mewes, H. W., Well, B. and
 Wiemann, S.
 TITLE EST (Ottenwaelder, B., Obermaier, B., Mewes, H. W., Well, B. and Wiemann
 S.)
 JOURNAL Unpublished (2001)
 COMMENT Contact: Ottenwaelder B
 MIPS Am Kioferplatz 18a D-82152 Martinsried, Germany
 This is the 5' sequence of the clone insert

Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
 Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;
 sequenced by Medigenomix (Martinsried/Germany) within the cDNA
 sequencing consortium of the German Genome Project. No sl sequence
 available.

This clone (DKFZ451L0110) is available at the RZPD in Berlin.
 Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
 Location/Qualifiers

FEATURES

source
 1. 376
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="DKFZ451L0110"
 /clone_1ib="451 (synonym: hlccl) spinal cord"
 /tissue_type="human spinal cord"
 /dev_stage="adult"
 /lab_host="DH10B"
 /note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI"
 BASE COUNT 127 a 63 c 79 g 107 t
 ORIGIN

Query Match 22.3%; Score 312; DB 9; Length 376;
 Best Local Similarity 100.0%; Pred. No. 2.8e-152; Length 376;
 Matches 312; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 550 aggacacagttctcagatgtcttgcattatctatctggaacccatttaactgataat 609
 DB 1 AGGACACAGTTCCTCAGATGTTCTGATCTTATCTGGAACCCATTTAAACCTGATAAT 60
 QY 610 tgcacatgttcagattctctctcttcgtgaatgcgaatgcaaaagtctggaatttcg 669
 |||||||
 DB 61 TGCAAAATGTTCAAGATTTCCTCTCTGTAATGCAAAATGCGAAAGTTTGGAAATTCG 120
 QY 670 tctgttcacagatcagacagaatgttccagaatcttttgacaacacgcctcagatgg 729
 |||||||
 DB 121 TCTGTTCCAGCATCAGACAGATGTCAGATCTTTTGGAACAACGCGCTCAGATGG 180
 QY 730 gcagaactctcaatgttcttcaatcacatggaataaattctatttgaataataga 789
 |||||||
 DB 181 GCAGAACTTCAATGTTCTTAATCACATGGAATAAATTTCTATTGTGAATAATGATAAG 240
 QY 790 agtctcagctcctaattgtcgtatcagtaaggcaaaatcacagatgtgagacaga 849
 |||||||
 DB 241 AGTACTGAGCTCTAAATGTCGTATTCAATAAGGCAAAATCACAGATGAGACAGA 300
 QY 850 ttgaaaaggaaa 861
 |||||||
 DB 301 TTGAAAAGGAAA 312

RESULT 3
 BM472954 1010 bp mRNA linear EST 05-FEB-2002
 LOCUS BM472954.6466106 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:5574588
 DEFINITION 5', mRNA sequence.
 ACCESSION BM472954
 VERSION BM472954.1 GI:18521996
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 1010)
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/
 TITLE NIH-MGC http://mgi.nci.nih.gov/
 JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
 COMMENT Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cga@bbs.femail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I M A G E Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLAM12323 row: h column: 13

High quality sequence stop: 738.

Location/Qualifiers

1. 1010

source

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:557458"

/clone_lib="NIH_MGC_88"

/tissue_type="duodenal adenocarcinoma, cell line"

/lab_host="DH10B (phage-resistant)"

/note="Organ: small intestine; Vector: pCMV-SPORT6;

Site_1: NotI; Site_2: SalI; Cloned unidirectionally;

oligo-dr primed. Average insert size 1.767 kb. Library

enriched for full-length clones and constructed by life

Technologies. Note: this is a NIH_MGC Library."

BASE COUNT 293 a 209 c 191 g 317 t

ORIGIN

Query Match

14.6%; Score 205; DB 10; Length 1010;

Best Local Similarity 99.5%; Pred. No. 3.8e-96;

Matches 375; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 920 cagttcaacaataaactactatgatacttgagagcttgaagaaattactactcgt 979

DB 20 CAGTTCAACAATAAATACCTATACATGATGATGAGGAGCTCTGAAGCAAACTTACTACGT 79

QY 980 ggttaatttgtaataagctggctttcaagatgtgagaagaatgataaagtgtcat 1039

DB 80 GGTAAATTGTATATAGTTCAGGCTTTCAAGTATGGAAGAATATCAAAAGTGTCTATAT 139

QY 1040 cagctcaagaagaacagatatactctcttccacagtggtgagtagctgcctgcatta 1099

DB 140 CAGCTCAACAAGAACAAGATATACCTTTTCCACAGCTGATGAGTGCCTGCACATTA 199

QY 1100 ctcaacaactt-aaaaacccatagggccttctgttttttgccttcttcttcttgagact 1158

DB 200 CTCACAACATTAATAAAAAACCCCATAGGCGCTTCGTTTTCCTCTTCTTGAGACAT 259

QY 1159 gggctgcattctatctgtgaatggccacttgcctgcctgattccagcagaagaacctgtg 1218

DB 260 GGGCTGCATTTTCATGGAATGGCCACTTTGCACCTGATTCACCAAGAACACTGTGG 319

QY 1219 cgtgatgataatgaggtgtgttcgaagtgcactggaataacagtttaagtacagcatta 1278

DB 320 CGTGATGATTAATGAGTGTGTTCGAAAGTGAAGTGAATTAAGTGAAGCATTA 379

QY 1279 atagctctgcataatgt 1295

DB 380 ATAGCTCCTCATATATGT 396

RESULT 4
AA776169 209 bp mRNA linear EST 05-FEB-1998

LOCUS AA776169 ae80c02.s1 StrataGene schizo brain S11 Homo sapiens cDNA clone

DEFINITION IMAGE:970466 3', mRNA sequence.

ACCESSION AA776169

VERSION AA776169.1 GI:2835503

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 209)

AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,

Kritman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Martin,M., Martin

J., Moore,B., Schellenberg,K., Stepien,M., Tan,F., Theising,B.,

White,Y., Wylie,T., Waterston,R. and Wilson,R.

TITLE WashU-NCI human EST Project

JOURNAL

Unpublished (1997)

Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@wustl.wustl.edu

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Possible reversed clone: polyT not found

Seq primer: -40ml3 fwd. ET from Amersham

High quality sequence stop: 206.

FEATURES

source

1. 209

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:970466"

/clone_lib="Stratagene schizo brain S11"

/sex="male"

/tissue_type="schizophrenic brain S-11 frontal lobe"

/dev_stage="34 years old"

/lab_host="SOLR (kanamycin resistant)"

/note="Vector: Bluescript SK-; Site_1: EcoRI; Library

constructed from S-11 frontal lobe, male, 34 years old,

50% caucasian, 50% Aleutian. Schizophrenic suicide.

Random primed into EcoRI site of ZAP II Vector. Mass

excised. Custom library. Avg insert length 1.4kb.

Material obtained by Johnston N., Torrey, E.F., Yolken R.,

and the Stanley Neuropathology Consortium - Analysis of

RNAs from the Brains of individuals with Psychiatric

Diseases (Unpublished) Stanley Neuropathology Laboratory,

Johns Hopkins School of Medicine, Baltimore MD."

BASE COUNT 58 a 42 c 47 g 62 t

ORIGIN

Query Match

14.3%; Score 201; DB 9; Length 209;

Best Local Similarity 100.0%; Pred. No. 5.1e-94;

Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 501 caactccattagctccggagacatggttagagattgagcagatttagacacagct 560

DB 9 CACATCCATTAGTCCGGAGCAGCATGTTAGATGTTGATGACAGATTTAGGACACAGTT 68

QY 561 ctcaagatgtctgagatcttatactggaacccatttaactggaataatgacacagct 620

DB 69 CTCGATGTTCTGTGATCTTATCTGGACACCATTTAACTGATTAATTGCACAAATGTT 128

QY 621 cagatttcctcttcgttaatgcagaatgcaaaagtttcggaaatctcgtcttcagc 680

DB 129 CAGATTTCCTCTTCGTAAATGCAGAAATGCGAAATGTCGATTCGTTCCAGC 188

QY 681 atcagacagaatggttcagaa 701

DB 189 ATCAGACAGATGTCAGAA 209

RESULT 5

R17106 795 bp mRNA linear EST 12-JUN-1996

LOCUS R17106 EST20108 Clontech adult human fat cell library HL1108A Homo sapiens

DEFINITION CDNA clone 20108, mRNA sequence.

ACCESSION R17106

VERSION R17106.1 GI:770716

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 795)

AUTHORS Boulland,F.

TITLE Study of expressed sequences tags in adipose tissue 1995

JOURNAL Unpublished (1995)

COMMENT Contact: Frederic Boullaud
Centre de Recherche sur l'Endocrinologie moléculaire et le
Développement
CNRS
9, Rue Jules Hetzel, Meudon Bellevue, 92190 France
Tel: 33 1 45 07 52 87
Fax: 33 1 45 07 58 90
Email: boullaud@infobiogen.fr
Southern human DNA EcoRI single band 2.9 Kb.
Location/Qualifiers
1. 795
/organism="Homo sapiens"
/strain="Caucasian"
/db_xref="taxon:9606"
/clone="20108"
/note="Vector: lambda gt10; Site 1: EcoRI; dev-stage=adult
; tissue-type=adipose tissue; lab_host=Bacteriophage
lambda; first strand priming with random and poly-d(T)
oligonucleotides."

FEATURES
source
1. 795
/organism="Homo sapiens"
/strain="Caucasian"
/db_xref="taxon:9606"
/clone="20108"
/note="Vector: lambda gt10; Site 1: EcoRI; dev-stage=adult
; tissue-type=adipose tissue; lab_host=Bacteriophage
lambda; first strand priming with random and poly-d(T)
oligonucleotides."

BASE COUNT 243 a 150 c 166 g 216 t 20 others
ORIGIN

Query Match 12.1%; Score 169; DB 10; Length 795;
Best Local Similarity 99.3%; Pred. No. 3e-77;
Matches 269; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 349 ggaagagccggaagaggaagaaatccttataaacctggaagatggaatgattcaat 408
|||||
Db 21 GGAAGAGCCGGAAGAGGGAATCCTTATTAAGATGGAAGATGGAATGATTCAT 80
|||||

QY 409 tccgtgatcatatccagagctgccatcttattcttgcaatgacatccgtgatt 468
|||||

Db 81 TCTGTGATATATTCACAGACTGCCCATCTTTATTCTGCAATGACATCTGTGATT 140
|||||

QY 469 ttgatctccatccagatatgacacagggacacatcatagtcctccgacgcatgtt 528
|||||

Db 141 TTGATATCTCATGCGAGATATGCACAGGGCCCATCATCTAGTCCCGGACCATGTT 200
|||||

QY 529 aagagattgagatcagatttagaacacagtttcagatgttcttgatcttctggga 588
|||||

Db 201 AGAGATTGGATGAGATTTTAGACACAGNTCTCAGATGTTCTGATCTTTATCTGGA 260
|||||

QY 589 acccatltaactgataatgcacatgt 619
|||||

Db 261 ACCCATTTAACTGATATTCACACATGT 291
|||||

RESULT 6
AA776670 148 bp mRNA linear EST 05-FEB-1998
LOCUS ae80602.s1 Stratagene scilzo brain S11 Homo sapiens cDNA clone
DEFINITION IMAGE:970514 3', mRNA sequence.
ACCESSION AA776670
VERSION AA776670.1 GI:2836004
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 148)
AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, J., Geisel, G., Jost, S.,
Kritman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Maita, M., Martin
White, Y., Wylie, T., Waterston, R., and Wilson, R.
JOURNAL Unpublished (1997)
COMMENT Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

EMAIL: est@watson.wustl.edu
This clone is available royally-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Possible reversed clone: polyT not found
Seg primer: -40m3 fwd. ET from Amersham
High quality sequence stop: 131.
Location/Qualifiers
1. 148
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:970514"
/clone_lib="Stratagene scilzo brain S11"
/sex="male"
/tissue_type="schizophrenic brain S-11 frontal lobe"
/dev_stage="34 years old"
/note="Vector: Bluescript SK-; Site 1: EcoRI; library
constructed from S-11 frontal lobe, male, 34 years old,
50% caucasian, 50% Aleutian. Schizophrenic suicide.
Random primed into EcoRI site of ZAP II Vector. Mass
excised. Custom library. Avg insert length 1.4kb.
Material obtained by Johnston N., Torrey, E.F., Yolken R.,
and the Stanley Neuropathology Consortium - Analysis of
RNAs from the Brains of individuals with Psychiatric
Diseases (Unpublished) Stanley Neurovirology Laboratory,
Johns Hopkins School of Medicine, Baltimore MD."

FEATURES
source
1. 148
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:970514"
/clone_lib="Stratagene scilzo brain S11"
/sex="male"
/tissue_type="schizophrenic brain S-11 frontal lobe"
/dev_stage="34 years old"
/note="Vector: Bluescript SK-; Site 1: EcoRI; library
constructed from S-11 frontal lobe, male, 34 years old,
50% caucasian, 50% Aleutian. Schizophrenic suicide.
Random primed into EcoRI site of ZAP II Vector. Mass
excised. Custom library. Avg insert length 1.4kb.
Material obtained by Johnston N., Torrey, E.F., Yolken R.,
and the Stanley Neuropathology Consortium - Analysis of
RNAs from the Brains of individuals with Psychiatric
Diseases (Unpublished) Stanley Neurovirology Laboratory,
Johns Hopkins School of Medicine, Baltimore MD."

BASE COUNT 37 a 30 c 33 g 48 t
ORIGIN

Query Match 10.0%; Score 140; DB 9; Length 148;
Best Local Similarity 100.0%; Pred. No. 5.3e-62;
Matches 140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 501 cacatccatagatcccgagcagcatgtttgagattgagatgagatgagatgagacacagt 560
|||||

Db 9 CACATCCATTATGTCGCGAGCATGTTTATGAGATTGATGATGATGATGATGATGAT 68
|||||

QY 561 ctcaagatgtctgagatcttatctgggaaccccttaactgataatgcacatgtt 620
|||||

Db 69 CTCAGATGTTCTGATCTTATCTGGAACCCATTTAACTGATATTCACATGTT 128
|||||

QY 621 cgaatttcctctcgtaagt 640
|||||

Db 129 CAGATTTCCTCTGTAATG 148
|||||

RESULT 7
AA897178 238 bp mRNA linear EST 04-JAN-1999
LOCUS am09e08.s1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
DEFINITION IMAGE:1466342 3', mRNA sequence.
ACCESSION AA897178
VERSION AA897178.1 GI:3033798
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 238)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps@mail.nih.gov
This clone is available royally-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 847 Std Error: 0.00
Seg primer: -40m3 fwd. ET from Amersham
High quality sequence stop: 132.
Location/Qualifiers
1. 238

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1466342"
/clone_lib="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site 1: Not I; Site 2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NDHL19W, testis NHT, and B-cell
NCI-CGAP_GCB1) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo."
BASE COUNT      71 a      53 c      45 g      69 t
ORIGIN

Query Match      10.0%; Score 140; DB 9; Length 238;
Best Local Similarity 100.0%; Pred. No. 5.2e-62;
Matches 140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1156 actggcgtgcatttcattgtaagtcgaccttgcactggaattcagccagaagaacctg 1215
    |||||||
DB 238 ACTGGCGTGCATTTCATTGTAAGTCGACCTTGCACGATTCAGCCAGAACCACTG 179

QY 1216 tggcgatgataagtcagtgctgttcggaagtcagtaacagttatgacagca 1275
    |||||||
DB 178 TGGCGTATGATGATGAGTTGCTTCGAACTGAGATGGAATACAGTTTATGACAGCA 119

QY 1276 ttaatagctcctgcatatgt 1295
    |||||||
DB 118 TTAAATAGCTCCTGCATATGT 99

RESULT 8
BH157201/c      665 bp      DNA      linear      GSS 24-SEP-2001
LOCUS          EMTSL057F Entamoeba histolytica Sheared DNA Entamoeba histolytica
DEFINITION     genomic DNA sequence.
ACCESSION      BH157201
VERSION        BH157201.1 GI:15730639
KEYWORDS       GSS.
SOURCE         Entamoeba histolytica.
ORGANISM       Entamoeba histolytica.
REFERENCE      1 (bases 1 to 665)
AUTHORS        Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.
TITLE          Determination of clone end sequences from Entamoeba histolytica
JOURNAL        HM1:IMSS sheared DNA library (2001)
COMMENT        Unpublished (2001)
                Contact: Brendan J Loftus
                Department of Eukaryotic Genomics
                The Institute for Genomic Research
                9712 Medical Center Dr., Rockville, MD 20850, USA
                Tel: 301 838 0208
                Fax: 301 838 3543
                Email: b1loftus@tigr.org
                Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
                DNA library
                Seq primer: M13-Forward
                Class: Shotgun
                High quality sequence start: 18
                High quality sequence stop: 569.
FEATURES
SOURCE         location/Qualifiers
                1..665
                /organism="Entamoeba histolytica"
                /strain="HM1:IMSS"
                /db_xref="taxon:5759"
                /clone_lib="Entamoeba histolytica Sheared DNA"

```

```

/note="Vector: pHOSt; Site 1: Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + 1 method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. M. Vaubin and B.
Barrell, Oxford University Press, 1999)."
BASE COUNT      237 a      125 c      82 g      221 t
ORIGIN

Query Match      1.6%; Score 22; DB 12; Length 665;
Best Local Similarity 100.0%; Pred. No. 4.3;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 205 gaatcgtgttggttgatc 226
    |||||||
DB 28 GAAATCTGTTTGTCTTGATC 7

RESULT 9
AV214110      247 bp      mRNA      linear      EST 30-OCT-1999
LOCUS          AV214110 RIKEN full-length enriched, ES cells Mus musculus cDNA
DEFINITION     clone 2410133H03.3' similar to AF015309 Mus musculus nucleolar
                protein (MSP58) mRNA, mRNA sequence.
ACCESSION      AV214110
VERSION        AV214110.1 GI:6154956
KEYWORDS       EST.
SOURCE         house mouse.
ORGANISM       Mus musculus.
REFERENCE      1 (bases 1 to 247)
AUTHORS        Konno,H., Alzawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T.,
                Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F.,
                Ishii,Y., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai
                ,C., Kawai,T., Kikuchi,N., Kojima,Y., Koya,S., Kusabe,M.,
                Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,
                Owa,C., Ozawa,Y., Saito,H., Sano,M., Sato,K., Shibata,K., Shibata
                ,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugihara,Y., Suzuki,H.,
                Suzuki,H., Takahashi,F., Tateno,M., Tomioka,N., Tsunoda,Y.,
                Watabiki,A., Watanabe,S., Yamamura,T., Yasunishi,A., Yokota,T.,
                Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
                RIKEN Mouse ESTs (Konno,H., et al. 1999)
                Unpublished (1999)
JOURNAL        Laboratory for Genome Exploration Research Group, RIKEN Genomic
COMMENT        Sciences Center(GSC), Yokohama Institute
                The Institute of Physical and Chemical Research (RIKEN)
                1-7-22 Suenho-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan
                Tel: 81-45-503-9222
                Fax: 81-45-503-9216
                Email: genome-res@sc.riken.go.jp,
                URL: http://genome.gsc.riken.go.jp/
                Sasaki,N., Izawa,M., Watabiki,M., Ozawa,K., Tanaka,T., Yoneda,Y.,
                Matsura,S., Carninci,P., Muramatsu,M., Okazaki,Y. and Hayashizaki
                ,Y.
                Transcriptional sequencing: A method for DNA sequencing using RNA
                polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)
                Itoh,M., Katsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
                Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
                ,Y. and Hayashizaki,Y.
                Automated filtration-based high-throughput plasmid preparation
                system. Genome Res. 9 (5), 463-470 (1999)
                Carninci,P. and Hayashizaki,Y.
                High-efficiency full-length cDNA cloning. Methods Enzymol. 303,

```


OY 853 aaagaacaattcatgca 873
 |||||||
 Db 504 AAAGGAAACATTTCATGCA 524

RESULT 12
 B70904 758 bp DNA linear GSS 21-JUN-1998
 LOCUS B70904
 DEFINITION C17-HSP-2063E16.TF C17-HSP Homo sapiens genomic clone 2063E16, DNA
 sequence.
 ACCESSION B70904
 VERSION B70904.1 GI:2710128
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 758)
 Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Golden
 K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M.
 and Venter,J.C.
 Use of a random BAC End Sequence Database for Sequence-Ready Map
 Building
 Unpublished (1997)
 Other GSSs: C17-HSP-2063E16.TR
 CONTACT: Mark Adams
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: mdadams@tigr.org
 Clones are available from Research Genetics (info@resgen.com). BAC
 end search page:
 http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html
 Seq primer: M13-21
 Class: BAC ends.

FEATURES
 source Location/Qualifiers
 1..758
 /organism="Homo sapiens"
 /db_xref="GDB:7061491"
 /db_xref="taxon:9606"
 /clone="2063E16"
 /clone_1b="C17-HSP"
 /sex="Male"
 /cell_type="Sperm"
 /note="Vector: pBelobAC11; site_1: HindIII; site_2:
 HindIII"

BASE COUNT 259 a 133 c 123 g 243 t

ORIGIN

Query Match 1.5%; Score 21; DB 12; Length 758;
 Best Local Similarity 100.0%; Pred. No. 14;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1332 ggttcgtatccaacattatca 1352
 |||||||
 Db 736 GGTCGTGATCCACATATATCA 756

RESULT 13
 B624213 227 bp mRNA linear EST 01-JUL-2001
 LOCUS B624213
 DEFINITION CVEMB-140 CVEMB Crassostrea virginica cDNA, mRNA sequence.
 ACCESSION B624213
 VERSION B624213.1 GI:14580643
 KEYWORDS EST.
 SOURCE eastern oyster.
 ORGANISM Crassostrea virginica
 Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Ostreoida;
 Ostreidae; Ostreidae; Crassostrea.

REFERENCE 1 (bases 1 to 227)
 AUTHORS Gross,P.S. and Bartlett,T.C.
 TITLE Crassostrea virginica Embryo EST Library (CVEMB)
 JOURNAL Unpublished (2001)
 COMMENT Contact: Jenny MJ
 Marine Biomedicine and Environmental Sciences
 Medical University of South Carolina
 221 Port Johnson Road, Charleston, SC 29412, USA
 Tel: 843 876 5058
 Fax: 843 762 5530
 Email: jenny@emusc.edu.

FEATURES
 source Location/Qualifiers
 1..227
 /organism="Crassostrea virginica"
 /db_xref="taxon:9565"
 /clone_1b="CVEMB"
 /tissue_type="Embryo, D-veliger larvae"
 /note="Vector: pTribEx2; site_1: Sfi I; site_2: Sfi I"

BASE COUNT 71 a 40 c 44 g 72 t

ORIGIN

Query Match 1.4%; Score 20; DB 10; Length 227;
 Best Local Similarity 100.0%; Pred. No. 49;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1013 tggagaaagtatctaaagt 1032
 |||||||
 Db 48 TGCAGAAAGTATCTAAAGT 67

RESULT 14
 A1705955 308 bp mRNA linear EST 03-JUN-1999
 LOCUS A1705955
 DEFINITION UI-R-AC0-Y1-b-07-0-UI.s1 UI-R-AC0 Rattus norvegicus cDNA clone
 UI-R-AC0-Y1-b-07-0-UI 3', mRNA sequence.
 ACCESSION A1705955
 VERSION A1705955.1 GI:4993855
 KEYWORDS EST.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sclurognath; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 308)
 Bonaldo,M.F., Lennon,G. and Soares,M.B.
 Normalization and subtraction: two approaches to facilitate gene
 discovery
 Genome Res. 6 (9), 791-806 (1996)
 JOURNAL 97044477
 MEDLINE
 COMMENT Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: msoares@blue.weeg.uiowa.edu
 Oligo-dT track not found. Not 1 site shown in beginning of sequence
 is likely internal to the message. cDNA Library Preparation: M.B.
 Soares lab clone distribution: clones will be available through
 Research Genetics (www.resgen.com) The following repetitive
 elements were found in this cDNA sequence: 263-301, >(CAG
)n#Simple_repeat
 Seq primer: M13 Forward
 POLYA-No.

FEATURES
 source Location/Qualifiers
 1..308
 /organism="Rattus norvegicus"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone="UI-R-AC0-Y1-b-07-0-UI"
 /clone_1b="UI-R-AC0"
 /dev_stage="adult"

/lab_host="DH10B (Life Technologies)"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-AGO
 library is a non-normalized library constructed from 16.5
 dpc rat atriocentricular (AV) canal. The tag is a string
 of 5 nucleotides present between the Not I site and the
 oligo-dt track. The library was constructed as described
 by Bonaldo, Lennon and Soares, Genome Research 6: 791-806
 , 1996. Tissue provided by Jim Lin, Department of Biology,
 University of Iowa.
 TAG_SPO=None found"
 BASE COUNT 46 a 77 c 78 g 107 t
 ORIGIN

Query Match 1.4%; Score 20; DB 9; Length 308;
 Best Local Similarity 100.0%; Pred. No. 49;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1144 cttctcttggaactggct 1163
 ||||||||||||||||
 Db 7 CTTCTTTGGAGACTGGCT 26

RESULT 15
 AI579360 320 bp mRNA linear EST 05-APR-1999
 LOCUS UI-R-AGO-wu-h-08-0-UI.s1 UI-R-AGO Rattus norvegicus cDNA clone
 DEFINITION
 ACCESSION AI579360
 VERSION AI579360.1 GI:4563736
 KEYWORDS EST.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 320)
 Bonaldo,M.F., Lennon,G. and Soares,M.B.
 Normalization and subtraction: two approaches to facilitate gene
 discovery
 Genome Res. 6 (9), 791-806 (1996)
 97044477
 Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: msoares@blue.weeg.uiowa.edu
 Oligo-dt track not found. Not I site shown in beginning of sequence
 is likely internal to the message. cDNA library preparation: M.B.
 Soares Lab Clone distribution: clones will be available through
 Research Genetics (www.resgen.com) The following repetitive
 elements were found in this cDNA sequence: 264-317, >(CAG
)n\$imple_repeat
 Seq primer: M13 Forward.
 Location/Qualifiers
 1..320

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 COMMENT

FEATURES

source
 /organism="Rattus norvegicus"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone="UI-R-AGO-wu-h-08-0-UI"
 /clone.lib="UI-R-AGO"
 /dev_stage="adult"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-AGO
 library is a non-normalized library constructed from 13
 dpc rat ventricle. The tag is a string of 6 nucleotides
 present between the Not I site and the oligo-dt track.
 The library was constructed as described by Bonaldo,
 Lennon and Soares, Genome Research 6: 791-806, 1996.

Tissue provided by Jim Lin, Department of Biology,
 University of Iowa."
 BASE COUNT 47 a 80 c 80 g 113 t
 ORIGIN

Query Match 1.4%; Score 20; DB 9; Length 320;
 Best Local Similarity 100.0%; Pred. No. 49;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1144 cttctcttggaactggct 1163
 ||||||||||||||||
 Db 7 CTTCTTTGGAGACTGGCT 26

Search completed: May 22, 2002, 07:30:35
 Job time: 10805 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 22, 2002, 08:30:10 ; Search time 3328.52 Seconds
(without alignments)
8808.142 Million cell updates/sec

Title: US-09-693-205-7_COPY_5300_6700

Perfect score: 1401

Sequence: 1 gtacgagtaaacactagagc.....aaagacacttaagaagt 1401

Scoring table: OLIGO_NUC

Searched: 1797656 seqs, 10463268293 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl: *
1: gb_ba: *
2: gb_htg: *
3: gb_in: *
4: gb_cm: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_sts: *
12: gb_sy: *
13: gb_un: *
14: gb_vl: *
15: gb_ba: *
16: gb_fun: *
17: em_hum: *
18: em_in: *
19: em_mu: *
20: em_cm: *
21: em_or: *
22: em_ov: *
23: em_pat: *
24: em_ph: *
25: em_pl: *
26: em_ro: *
27: em_sts: *
28: em_un: *
29: em_vl: *
30: em_htg_hum: *
31: em_htg_inv: *
32: em_htg_other: *
33: em_hugo_inv: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result Query No. Score Match Length DB ID Description

1	1295	92.4	12793	6	AX119931	AX119931 Sequence	
2	1295	92.4	12793	9	AF193556	AF193556 Homo sapi	
3	1295	92.4	92693	9	AL157766	AL157766 Human DNA	
4	923	63.9	99819	2	AC079761	AC079761 Homo sapi	
5	53	3.8	11492	6	AX119933	AX119933 Sequence	
6	53	3.8	11493	10	AF193557	AF193557 Mus muscu	
7	23	1.6	189760	9	AC022025	AC022025 Homo sapi	
8	22	1.6	163869	9	AL356967	AL356967 Human DNA	
9	21	1.5	2296	2	AC013913	AC013913 Drosophila	
10	21	1.5	55954	2	AL136322	AL136322 Human DNA	
11	21	1.5	110000	2	AC091528_1	Continuation (2 of	
12	21	1.5	129504	9	AL445469	AL445469 Human DNA	
13	21	1.5	157913	2	AC106679	AC106679 Rattus no	
14	21	1.5	171926	2	AC084374	AC084374 Homo sapi	
15	21	1.5	176912	2	AC096964	AC096964 Rattus no	
16	21	1.5	212888	9	AL513325	AL513325 Human DNA	
17	21	1.5	262059	3	AE003573	AE003573 Drosophila	
18	21	1.5	303040	1	AP000991	AP000991 Thermopila	
19	20	1.4	624	10	MMVH81XR2	U16744 Mus musculu	
20	20	1.4	1208	8	AY010964	AY010964 Hordeum v	
21	20	1.4	1218	8	AY010965	AY010965 Elymus el	
22	20	1.4	1218	8	AY010966	AY010966 Elymus g1	
23	20	1.4	1218	8	AY010967	AY010967 Elymus g1	
24	20	1.4	1218	8	AY010968	AY010968 Elymus g1	
25	20	1.4	1218	8	AY010971	AY010971 Elymus r1	
26	20	1.4	1218	8	AY010975	AY010975 Elymus v1	
27	20	1.4	1218	8	AY010976	AY010976 Elymus v1	
28	20	1.4	1218	8	AY010977	AY010977 Elymus wa	
29	20	1.4	1218	8	AY010978	AY010978 Elymus wa	
30	20	1.4	1219	8	AY010961	AY010961 Hordeum b	
31	20	1.4	1242	8	AY010981	AY010981 Elymus g1	
32	20	1.4	1247	8	AF331953	AF331953 Elymus v1	
33	20	1.4	3235	10	RATNOP140B	M94288 Rattus norv	
34	35	20	1.4	3609	10	RATNOP140A	M94287 Rattus norv
35	20	1.4	10310	10	AC084885	AC084885 CgFancG/X	
36	20	1.4	33758	3	CEM79	250806 Caenorhabdi	
37	20	1.4	36823	9	HSU20961	268873 Human DNA s	
38	20	1.4	40142	2	AC015137	AC015137 Drosophila	
39	20	1.4	49438	2	AC106585	AC106585 Rattus no	
40	20	1.4	72207	9	AL596254	AL596254 Human DNA	
41	20	1.4	74328	2	AC100878	AC100878 Mus muscu	
42	20	1.4	102104	2	AC106430	AC106430 Rattus no	
43	20	1.4	121067	2	AC096119	AC096119 Rattus no	
44	20	1.4	135499	9	AC026446	AC026446 Homo sapi	
45	20	1.4	139887	9	CNS01RGX	AL160233 Human chr	

ALIGNMENTS

RESULT 1
AX119931
LOCUS AX119931 12793 bp DNA linear PAT 11-MAY-2001
DEFINITION Sequence 1 from Patent WO0129266.
ACCESSION AX119931
VERSION AX119931.1 GI:14036678
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 12793)
AUTHORS Hudson,T.J., Engert,J. and Richter,A.
TITLE Identification of arsacs mutations and methods of use therefor
JOURNAL Patent: WO 0129266-A 1 26-APR-2001;
MCGILL UNIVERSITY (CA) ; Hopital Sainte-Justine (CA)

FEATURES
source
1..12793
Location/Qualifiers

BASE COUNT 4163 a 2256 c 2487 g 3887 t
ORIGIN
/organism="Homo sapiens"
/db_xref="taxon:9606"

Query Match 92.4%; Score 1295; DB 6; Length 12793;
 Best Local Similarity 100.0%; Pred No. 0;
 Matches 1295; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 gtagcagtaaaactagagcagtcaccaagcgacacaaagccttagaagaatagatgcc 60
    |||||
DB 5300 GTAGCAGTAAACTAGAGAGAGTCCCAAGCGACACAAAGCCTTAGAAGATATGCAATCC 5359
QY 61 aatgctctgttttaacaactctggcagacagaattgggcagaagaagaataatggaccagaga 120
    |||||
DB 5360 AATGCTGTGTTTACAACACTTGGCAGAGAAATTTGGCGAGAAAGAAATGGACACAGAGA 5419
QY 121 attaagagcactccttaataatgatacctctgaaaggaaatgttgaaggactcttcaa 180
    |||||
DB 5420 ATTAAAGCATCTCTTAATGATATCTCTTGGAAAGCAATGTGTAAGAGCTTTCTTCAA 5479
QY 181 aatgctgaatgtagcaaaagcgacagaatctgttctgttctgtagctagacagatcca 240
    |||||
DB 5480 AATGCTGATGATGCAAAAGGCGACAGAAATCTGTTTGTGTTGATCCTAGACAGCATCCA 5539
QY 241 gttgataagaatatttgatagaatgagtgcccaattgcaaggccagcaacttgtgtgtac 300
    |||||
DB 5540 GTTGAATAGATATTTGATGATTAAGTGGGCCCATTTGCAAGGCGCACACTTGTGTGTAC 5599
QY 301 aacaaccagcactttacagagaatgtagttagagaatcagaatcttggaaaggcagcag 360
    |||||
DB 5600 AACACACAGCATTTTACAGAGATGATGTTAGAGAAATTCAGAAATCTTGGAAAAAGCACGC 5659
QY 361 aaagagggaatccctataaaactgagcagtagagaatagatccaattctgtgtacat 420
    |||||
DB 5660 AAAGAGGGAATCTTATAAACTGAGACAGTATGAAATAGATGCAATCTGTGTATCAT 5719
QY 421 atacagcactggccactcttattcttctggaatagacatcctgtgtatttttgatcctcat 480
    |||||
DB 5720 ATACAGACAGTCCCATCTTTATTTCTGGCAATGACATCTGTGTATTTTGTGATCCATCAT 5779
QY 481 gccagatctgacacaggggacacatccatagtcacggacagcagctgtttagagaattggat 540
    |||||
DB 5780 GCCAGATATGACACAGGCGGCACATCATTAAGTCCCGAGCGCATGTTTAGAGATTTTGAT 5839
QY 541 gcagattttaagacacagttctcagatgltctgagatccttatcttgggaaccatlttaaa 600
    |||||
DB 5840 GCAGATTTTAAAGACACAGTCTCAGATGTTCTGATCTTATCTGGGAACCCATTTTAA 5899
QY 601 ctgagtaattgacacaatgltcagatctcctctcgtatgacaaatggaaagaatttgcg 660
    |||||
DB 5900 CTGGAATAAATGGCAAAATGTTCAAGATTTCTCTTGTATGCAAGAAATGCGAAAAAGTTTCG 5959
QY 661 gaaatttcgtctgttcacagatcagacaagaatgltccagaaatcttggacaacatgcgc 720
    |||||
DB 5960 GAAATTTTCGTCTGTTCACAGATCAGACAGATGTCAGAAATCTTTTGGCAAAATGCGCG 6019
QY 721 tcsagatggggcagaaactcttaattgttcttaacatcagaaaaaattctatttltgaa 780
    |||||
DB 6020 TCAGATGGGCGCAAGACTTCTAATGTTTCTTAATCACATGAAAAAATTTCTATTGTGAA 6079
QY 781 atagataaagatgactctgagctctaaatgtgtctgtattcgtlaaaggcaaaatcacagat 840
    |||||
DB 6080 ATAGATAAAGATGACTGAGCTCTAATGTCGTATTATTCATTAAGGCAAAATCACAGAT 6139
QY 841 gggagacagattgaaaaggaaacaattcatctgactctgtaattgatatgttactaaaga 900
    |||||
DB 6140 GGGAGACAGATTGAAAAGGAAACAATTTTCATGCTCTGTAATTTATGTTACTTAAAAAAG 6199
QY 901 aggcagctcaaaagacaatacagttcaacaataactatactatgatactgagagactc 960
    |||||
DB 6200 AGGCAGCTCAAAAGACATATACAGTTCAACAAATACCTATCTATGATGATCTAGGAGCTCT 6259
QY 961 gaagaaatcttactacgtggtcaatttgttaataagatcagggcttccaagtatgsgaaa 1020
    |||||
DB 6260 GAAGAAATCTTACTACGTGGCTAATTTGTAAATAGATCAGGCTTTTCAAGTATGAGAGAA 6319
  
```

```

QY 1021 gtactcaaaagtgtcatatcagctcacaagaaccagaatatactcttccacgtgt 1080
    |||||
DB 6320 GTATCTRAAAGTGTCTATATAGCTCACAGACACAGATTTACTCTTTCCACGTGT 6379
QY 1081 ggaagtagctgctgtcatctactcaacaactataaaacccatagggccttctgtt 1140
    |||||
DB 6380 GGAGTAGCTCTGTCAATTACTACAACTATPAAAAAACCCTATGAGGCTCTCTTTT 6439
QY 1141 cctcttcttlttgagagactgggcctgcatcttcaatgtagaagggcacttggactgttca 1200
    |||||
DB 6440 CTTCTTTCTTTTGGAGACTGGGCTGCTGCTTTCATGTAAGTGCACCTTGATGATCA 6499
QY 1201 gccagaagaaactgtgagcgtgtagaataatggagtggtgtctcgaagtgcagaaatc 1260
    |||||
DB 6500 GCCAGAAAGAACTGTGTGCTGATGATTAAGAGATTTGTTTGAAGTGTGATGATGAT 6559
QY 1261 agttaatgacagcaatlaagctcctgtcatatgt 1295
    |||||
DB 6560 AGTTAATGACACATTAATAGCTCCTGCATATGT 6594

RESULT 2
AF193556 12793 bp DNA linear PRI 07-FEB-2000
LOCUS AF193556 Homo sapiens saccin (SACS) gene, complete cds.
DEFINITION AF193556
ACCESSION AF193556
VERSION AF193556.1 GI:6907041
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 12793)
Engert,J.C., Berube,P., Mercier,J., Dore,C., Lepage,P., Ge,B.,
Bouchard,J.P., Machieu,J., Melancon,S.B., Schalling,M.,
Lander,E.S., Morgan,K., Hudson,T.J. and Richter,A.
ARSACS, a spastic ataxia common in northeastern Quebec, is caused
by mutations in a new gene encoding an 11.5-kb ORF
Nat. Genet. 24 (2), 120-125 (2000)
JOURNAL 20120709
MEDLINE 2 (bases 1 to 12793)
Engert,J.C., Berube,P., Dore,C., Lepage,P., Ge,B., Hudson,T.J. and
Richter,A.
Direct Submission
TITLE Submitted (08-Oct-1999) Genome Centre, Montreal General Hospital,
1650 Cedar Ave., Montreal, QC H3G 1M4, Canada
FEATURES
source
1..12793
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="13"
/map="between D13S232 and D13S292"
1..12793
/gene="SACS"
/product="sacsin"
1..12793
/gene="SACS"
77..11566
/gene="SACS"
/note="molecular chaperone"
/codon_start=1
/product="sacsin"
/protein_id="AF193556.1"
/db_xref="GI:6907042"
/translacion="MNTFEPGRELIVQWYTPDENRNHPSVSLKAWKNLYTHSEDL
TLFDEMPPLPTLLEGGQVCLIRLRPSLVILDDSEAOQLPEFLADIVOLGQVFL
KRLDASIQHPLIKKYLHSPVPAVQIMKVLQKCNQITSLPTHDLARKFLASL
TPSSSEKRIIOELAIKRIINSSDQGISYKLGCKVLTHTALPADLRISVID
SSDEAFIRANMLKTEOLKTSCLTLYIKDINAFYSHEVYOLMLVYENISLKE
NPNVLEMLPLKFTIOISOBQVYSAQELPEPDIKDLFCNEEGTYPFPYSVTSPTL
HSLRQIGLKNESLKEKDVQYAKIEALQVACPDQVLLKRAKATLLVLRKNTPL
OSSEKMTLKKIKVWPACKERPYPNGSLVWVGDCNLCAPDMKDVGHAILIGSLP
LVESIHVNLERKALGIFTRPSLSAVLKHFKITVVDWYSKTFSDQDYFOHILLETYGF
  
```


Qy 1201 gccagaagacacctgctgctgataatgatgctgtgttcgaagtgcactgaataac 1260
|||||
Db 6500 gccagaagaaacctgtgctgctgataatgatgctgtgttcgaactgactgaataac 6559
Qy 1261 agcttaatagcagcattaatagctctgcatatgt 1295
|||||
Db 6560 agcttaatagcagcattaatagctctgcatatgt 6594
RESULT 3
AL157766/c 92693 bp DNA linear PRI 11-APR-2001
LOCUS
DEFINITION Human DNA sequence from clone Rpl1-40020 on chromosome
13q12.11-12.2, complete sequence.
ACCESSION AL157766
VERSION AL157766.9 GI:13620292
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
AUTHORS Troman, A.
TITLE Direct Submission
JOURNAL Submitted (11-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Apr 12, 2001 this sequence version replaced gi:12709868.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: EMBL, SW, SWISSPROT, TrEMBL, WPI, WormPEP; information on the WormPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 13, constructed by the Sanger Centre Chromosome 13
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr13
Rpl1-40020 is from the library RPCI-11.1 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6
IMPORTANT: This sequence is not the entire insert of clone
Rpl1-40020. It may be shorter because we sequence overlapping
sections only once, except for a 100 base overlap.
The true left end of clone Rpl1-760M1 is at 92594 in this sequence.
The true right end of clone Rpl1-72P19 is at 100 in this sequence.

FEATURES
source
1..92693
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="13"
/map="q12.11-12.2"
/clone="Rpl1-40020"
/clone_lib="RPCI-11.1"
2390..2485
/note="MIR repeat: matches 81..192 of consensus"
2562..2673
/note="MIR repeat: matches 68..183 of consensus"
3896..4201
/note="AluY repeat: matches 3..308 of consensus"

repeat_region 5122..5397
/note="MER46c repeat: matches 1..286 of consensus"
18986..19294
/note="AluY repeat: matches 1..310 of consensus"
repeat_region 19644..19873
/note="MER46a repeat: matches 1..235 of consensus"
20613..20912
/note="AluSg1 repeat: matches 1..300 of consensus"
23342..23651
/note="AluSg1 repeat: matches 1..309 of consensus"
24769..24891
/note="L2 repeat: matches 2554..2662 of consensus"
25871..26011
/note="L2 repeat: matches 2356..2495 of consensus"
26033..26109
/note="L2 repeat: matches 2601..2668 of consensus"
26245..26344
/note="L2 repeat: matches 2154..2255 of consensus"
26938..27096
/note="MIR repeat: matches 3..175 of consensus"
27150..27653
/note="L2 repeat: matches 1063..1644 of consensus"
28522..28891
/note="THE1B repeat: matches 1..364 of consensus"
29447..29834
/note="L1ME3a repeat: matches 5787..6164 of consensus"
30098..36415
/note="AluSx repeat: matches 1..308 of consensus"
37202..37414
/note="MIR repeat: matches 22..262 of consensus"
37963..38254
/note="AluSg repeat: matches 9..301 of consensus"
38703..39008
/note="AluSg repeat: matches 1..306 of consensus"
39790..40093
/note="AluSx repeat: matches 1..304 of consensus"
40126..40416
/note="AluSg repeat: matches 1..292 of consensus"
40444..40733
/note="AluSg repeat: matches 1..292 of consensus"
41322..41405
/note="Single clone region. Assembly confirmed by
restriction digest data."
41541..41788
/note="AluSg repeat: matches 1..248 of consensus"
44790..45101
/note="AluSg repeat: matches 1..313 of consensus"
45261..45312
/note="13 copies 4 mer tggc 88% conserved"
45899..46206
/note="AluY repeat: matches 1..307 of consensus"
46754..47052
/note="AluY repeat: matches 1..298 of consensus"
47067..47365
/note="AluY repeat: matches 1..299 of consensus"
47477..47873
/note="L1MA10 repeat: matches 5950..6322 of consensus"
47889..48229
/note="AluSx repeat: matches 1..312 of consensus"
49168..49212
/note="Alu repeat: matches 85..126 of consensus"
49620..49693
/note="L2 repeat: matches 1685..1757 of consensus"
50704..51032
/note="AluSc repeat: matches 1..308 of consensus"
52204..53009
/note="CpG island"
53978..54137
/evidence="not_experimental"
/note="AluSx repeat: matches 1..160 of consensus"
54179..54511
/note="L1MB6 repeat: matches 5822..6172 of consensus"
55685..55949

repeat_region /note="AluX repeat: matches 1. .305 of consensus"
57331. .57390
/note="30 copies 2 mer ga 75% conserved"
repeat_region 57357. .57392
/note="9 copies 4 mer gaga 91% conserved"
repeat_region 57753. .57930
/note="MIR repeat: matches 82. .262 of consensus"
repeat_region 58260. .58389
/note="MIR repeat: matches 2. .153 of consensus"
repeat_region 58564. .58611
/note="24 copies 2 mer ca 93% conserved"
repeat_region 59350. .59533
/note="AluX repeat: matches 129. .313 of consensus"
repeat_region 59992. .60223
/note="AluX repeat: matches 85. .299 of consensus"
repeat_region 61036. .61144
/note="L2 repeat: matches 2581. .2696 of consensus"
repeat_region 62008. .62187
/note="TIGER1 repeat: matches 2238. .2418 of consensus"
repeat_region 62188. .62316
/note="AluX repeat: matches 1. .129 of consensus"
repeat_region 62330. .62363
/note="Alu repeat: matches 261. .294 of consensus"
repeat_region 62362. .62565
/note="TIGER1 repeat: matches 1586. .1787 of consensus"
repeat_region 62566. .62865
/note="AluX repeat: matches 1. .302 of consensus"
repeat_region 62866. .64385
/note="TIGER1 repeat: matches 46. .1586 of consensus"
repeat_region 64386. .64694
/note="AluX repeat: matches 1. .306 of consensus"
repeat_region 64695. .64713
/note="TIGER1 repeat: matches 29. .46 of consensus"
repeat_region 65068. .65395
/note="LIPB2 repeat: matches 5405. .5733 of consensus"
repeat_region 65396. .65569
/note="AluX repeat: matches 136. .309 of consensus"
repeat_region 65571. .65640
/note="LIPB2 repeat: matches 5728. .5791 of consensus"
repeat_region 65696. .65717
/note="11 copies 2 mer ta 100% conserved"
repeat_region 65725. .66096
/note="LIPB2 repeat: matches 5789. .6155 of consensus"
repeat_region 66371. .66410
/note="10 copies 4 mer tgtg 82% conserved"
repeat_region 67586. .67886
/note="AluX repeat: matches 1. .299 of consensus"
repeat_region 69748. .69930
/note="MIR repeat: matches 6. .248 of consensus"
repeat_region 70957. .71267
/note="AluX repeat: matches 1. .311 of consensus"
repeat_region 71279. .71413
/note="MER21B repeat: matches 548. .680 of consensus"
repeat_region 71411. .71737
/note="MER31A repeat: matches 47. .485 of consensus"
repeat_region 71780. .72075
/note="AluX repeat: matches 1. .295 of consensus"
repeat_region 72145. .72256
/note="MER31-internal repeat: matches 42. .175 of consensus"
repeat_region 72454. .72865
/note="MER31-internal repeat: matches 332. .739 of consensus"
repeat_region 72873. .73249
/note="MER31-internal repeat: matches 883. .1261 of

Query Match 92.4%; Score 1295; DB 9; Length 92693;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1295; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtgacagtaaaactagagcagtcacaaagcagacacaaagccttagaataatgcatcc 60
DB 13610 GTAGCAGTAAAGCTAGAGCAGTCCTCCAAAGCGACACAAAGCCTTGAAGAATATGATCCTC 13551

QY 61 aatgctcttttacacacttggcacagaatttggcagaagaataatgacacaga 120
DB 13550 AATGCTCTTTTACACACTTGGCACAGAAATTTGGCAGAAAGAAATTTACACAGCAGA 13491
QY 121 ataaagcaltccttaatgcalatcctctgtaaaaggaaatgttgaaagacttctcaa 180
DB 13490 ATTAAGAGCATCCTTAATGCAATATCTCTGAAAGAAATGTGAAAGACTTCTTCAA 13431
QY 181 aatgtctatgtcgaaggcgacagaatctgtttgttggcttcttagacacatcca 240
DB 13430 AATGCTATGATGCAAGGCGACAGAAATCTTTGTGTGATCTAGACACATCCA 13371
QY 241 gttagataaatatttgatgaagtgggcccatctgcaaggcgacacacttctgttac 300
DB 13370 GTTATGAATATTTGATGATAGTGGCCCATTTGCAAGGCGCCTTGTGTGATC 13311
QY 301 aacacacagccattacagaagaatgcatgtagaagaattcgaatcttgaaaaagcacg 360
DB 13310 AACACACAGCATTTACAGAAAGATGATGATGAGATTCGAAATCTTGAAAGGCGACG 13251
QY 361 aaagaaggaaatccttaataaaactgacagatggaataagatccaattctgtatcat 420
DB 13250 AAAGAGGAAATCCTTATAAACTGACAGATGGAATAGGATTCAAATTCGTATCAT 13191
QY 421 atcacagatgcccatcttatttcttgcgaatgacatccgtgtattttgaccccat 480
DB 13190 ATCACAGATGCCCATCTTTATTTTGTGGCAATGACATCCGTATTTTGTGATCCAT 13131
QY 481 gccagatatgacacaggggccacatccatagtcgccgacgacgtttagaatgtgat 540
DB 13130 GCCAGATATGACACAGGGGCCACATCCATTTATGTCGCCGACCATGTTAGATTTGAT 13071
QY 541 gcgaattttaggacacagttctcagatgttctggaatccttaactggaaccatttaaa 600
DB 13070 GCAATTTTGGACACAGTCTCAGATGTTCTGATCTTATCTGGGAAACCATTTTAAA 13011
QY 601 ctggaataatgacaaatgttgcagatcttctcttctgtaatgagaagaatggcaagttcg 660
DB 13010 CTGGAATATTTGCACAAATGTTGCAATTTCTCTTGTGTAATGCAAAAGGCAAAAGTTTCG 12951
QY 661 gaaattcgtctgttccagcatcagacagaatgttccagaatcttcttgacaaactgcgc 720
DB 12950 GAAATTTCTGTGTCTCCAGCATGACAGATGTCAGATCTTTTGGCAAACTCGCC 12891
QY 721 tcagatggggcagaactcttaatgttctcctaatacaatggaataaattctattgtgaa 780
DB 12890 TCAGATGGGCGAGAACTTCTTAATGTTTCTTAATCACAATGCAAAATTTCTAATTTGTGAA 12831
QY 781 atagataagtaactgagagctctaagtgtgtatctcagtaaaaggcaaatatcagat 840
DB 12830 ATAGATTAAGATGAGTGAAGCTTAATGCTGTAATGCTGTAATGCTGTAATGCTGTAATGCT 12771
QY 841 ggaagacagatgtaaaagaaacaattcatgcatctgtaattgtatgtatgtaactaaag 900
DB 12770 GGAACAGATTTGAAGAAGAAACAATTTCAATCACTGTAATGTATGATGTTACTAAAAAG 12711
QY 901 aggcagctcaaaagacataccagttcaacaataaactatactatgatactgagactct 960
DB 12710 AGGCAGCTCAAAAGACATNACGTTCAACAAATACTATACTATGATGATGAGAGACTCT 12651
QY 961 gaaagaaatcttaactagtgtaagtgttaagtcaagcttcttaagtaagtaagaa 1020
DB 12650 GAAAGAAATCTTACTAGCTGCTGAATTTGTAATGATGAGCTTTTCAAGTATGAGAGAA 12591
QY 1021 gtatctaaagtgtcatatcagctcacaagaacagatatctcttccacgtgtgt 1080
DB 12590 GTATCTAAAGTGTGATATATACGCTCACAGAACAGATATTTACTCTTTCCACGTGAT 12531
QY 1081 ggagtagctgtcgtcatatcactacaactataaaaaaccatagggccttctgtttttg 1140
DB 12530 GGAGTAGCTGCTGCTATCTACTACACAACTATTAATAAACCCCATFAGGCGCTTCTGTTTTTGG 12471

misc_feature 2639..3976
/note="assembly_name:Contig30"
misc_feature 4077..5355
/note="assembly_name:Contig35"
misc_feature 5456..6757
/note="assembly_name:Contig39"
misc_feature 6858..8570
/note="assembly_name:Contig40"
misc_feature 8671..9925
/note="assembly_name:Contig41"
misc_feature 10026..11426
/note="assembly_name:Contig42"
misc_feature 11527..13266
/note="assembly_name:Contig44"
misc_feature 13367..14794
/note="assembly_name:Contig47"
misc_feature 14895..16054
/note="assembly_name:Contig48"
misc_feature 16155..17395
/note="assembly_name:Contig49"
misc_feature 17496..19287
/note="assembly_name:Contig50"
misc_feature 19388..21294
/note="assembly_name:Contig51"
misc_feature 21395..22944
/note="assembly_name:Contig52"
misc_feature 23045..24421
/note="assembly_name:Contig53"
misc_feature 24522..25870
/note="assembly_name:Contig54"
misc_feature 25971..27230
/note="assembly_name:Contig55"
misc_feature 27331..28778
/note="assembly_name:Contig57"
misc_feature 28879..30893
/note="assembly_name:Contig58"
misc_feature 30994..32460
/note="assembly_name:Contig59"
misc_feature 32561..33984
/note="assembly_name:Contig60"
misc_feature 34085..35285
/note="assembly_name:Contig61"
misc_feature 35386..37184
/note="assembly_name:Contig62"
misc_feature 37285..39172
/note="assembly_name:Contig63"
misc_feature 39273..40874
/note="assembly_name:Contig64"
misc_feature 40975..42893
/note="assembly_name:Contig65"
misc_feature 42994..44384
/note="assembly_name:Contig66"
misc_feature 44485..45999
/note="assembly_name:Contig67"
misc_feature 46100..48669
/note="assembly_name:Contig68"
misc_feature 48770..50798
/note="assembly_name:Contig69"
misc_feature 50899..52809
/note="assembly_name:Contig70"
misc_feature 52910..55127
/note="assembly_name:Contig71"
misc_feature 55228..58087
/note="assembly_name:Contig72"
misc_feature 58188..61004
/note="assembly_name:Contig73"
misc_feature 61105..64185
/note="assembly_name:Contig74"
misc_feature 64286..67105

Query Match 65.9%: Score 923; DB 2; Length 99819;
Best Local Similarity 100.0%: Pred. No. 0;
Matches 923; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtagcagcraaacctcagagcagtcgccaaagcgacacaaagccttagaagaatagcatcc 60
DB 88196 GTAGCAGTAAACTAGAGGAGTCCCAAAAGCGACCAAAAGCCTTAGAAAGATATGCATCC 88137
QY 61 aatgtcttttaaacctctggcagcagaatttggcgagaagaataatcaccagaga 120
DB 88136 AATGCTCTTTTACACACTTGGCACACAGATTGGCGAAGAAATTTGCCACGACGA 88077
QY 121 ataaagacatccttaatgcatatccctcttgaagaagaatgttgaagaagccttctcaa 180
DB 88076 ATTAAGACATCTCTTAATGCAATTCCTTCGAAAAGAAATGTTGAAGAGCTTCTTCAA 88017
QY 181 aatgtcgtatgcagaagcgacagaatctgtttgtttgtttatccagaacatcca 240
DB 88016 AATGCTGTATGATGCAGAAAGCGACAGAAATCTGTTTGTGTTGAATCCAGACGATCA 87957
QY 241 gttgataagaatattgtagtaagtggcccatctggcagaagggccagcacttgtgtac 300
DB 87956 GTTGATGAATATTTGATGATGATGAGTGGCCCATTCGAAAGGCCAGCACTTGTGTATC 87897
QY 301 aacacccagcatctacagaagaatgtagtgaagaaatcttgaaagagcagc 360
DB 87896 AACAAACGACGATTTACAGAAAGATGATGTTAGAGAAATTCAGAAATCTTGAAAGACG 87837
QY 361 aaagagggaaatccttaataaacctggacagatagtaagatcaatctcgtatcat 420
DB 87836 AAAGAGGGAATCCCTTATTAATACTGGACATGATGGAATAGGATCAATTCGTGTATCAT 87777
QY 421 atcacagctgccatcttattcttctgcaatgacatccgtgattttagatccat 480
DB 87776 ATCAGAGCTGCCATCTTTATTTCTGCAATGACATCTGTTATTTTGTATCTCTCAT 87717
QY 481 gccagatatgcacagagggccacatccatagtcgccgagcagatgttagaatgtgat 540
DB 87716 GCCAGATATGACACGAGGGCCACATCCATGATGCCGACGACATGTTAGAGATTTGGAT 87657
QY 541 gcagattttagacaacagttctcagatgttctgatacttctcgggaaacccatttaa 600
DB 87656 GCAGATTTTAGACACACATCTTCAGATGTTCTGATCTTTATCTGGGAACCATTTTAAA 87597
QY 601 ctgataatgcacaatgttcagatttctctctctgtaatgagaagtgcaaaagtttcg 660
DB 87596 CTGATTAATTCACACATGTTTACATTTCTTCTCTTAATGCAAGAAATGCAAAAGTTTCG 87537
QY 661 gaaattcgtctgttcacagcatcagacagaaatgttcagaaatctttagacaactgcgc 720
DB 87536 GAAATTTGCTGCTCCAGCATCAGACAGAAATGTCAGAAATCTTTGGACAAACTGGCG 87477
QY 721 tcagatgggacagaactcttaatgttcttaatcaatgagaataatttcaattgtgaa 780
DB 87476 TCAGATGGGCGAGAACTTCAATGTTTCTTAATCAATGGAATAATTTCTAATTTGTAA 87417
QY 781 atagataagatctctggcgcttaaatgttgtatataagaaaggaaataatcaaat 840
DB 87416 ATAGATAAGATCTGGCGCTTAATGTGCTTAATGTAAGTAAGGCAAAATTCACACAT 87357
QY 841 ggaagacaatgtgaaagaaacaaatctcatgcatctgtaattgtagttagtactaaag 900
DB 87356 GGAGACATTTGAAGAAAGAAACAAATTTTCATGCTGTGAATTAATGATGCTTAATAAG 87297
QY 901 aggcagctcaagaacatccagc 923
DB 87296 AGGCAGCTCAAGACATCCAGT 87274

RESULT 5
AX119933 AX119933 11492 bp DNA linear PAT 11-MAY-2001
LOCUS
DEFINITION Sequence 3 from Patent WO0129266.
ACCESSION AX119933
VERSION AX119933.1 GI:14036679
KEYWORDS

SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS 1 (bases 1 to 11492)
TITLE Hudson, T.J., Engert, U. and Richter, A.
JOURNAL Identification of arasacs mutations and methods of use therefor
PATENT: WO 0129266-A 3 26-APR-2001;
MCGILL UNIVERSITY (CA); Hopital Sainte-Justine (CA)
FEATURES
source
1. 11492
/organism="Mus musculus"
/db_xref="taxon:10090"
BASE COUNT 3599 a 2280 c 2387 g 3226 t
ORIGIN
Query Match 3.8%; Score 53; DB 6; Length 11492;
Best Local Similarity 100.0%; Pred. No. 8e-17;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 298 tacacaacaccagcatttacagaagatgtagaggaattcaagaatcttg 350
|||||
Db 5520 TACAACACACGACCATTTACAGAAGATGATGTTAGAGGAATTCGATCTTGG 5572
|||||
RESULT 6
AF193557 11493 bp DNA linear ROD 07-FEB-2000
LOCUS Mus musculus sacsin gene, complete cds.
DEFINITION AF193557
ACCESSION AF193557 GI:6907043
VERSION AF193557.1
KEYWORDS house mouse.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 11493)
AUTHORS Engert, J.C., Berube, P., Mercier, J., Dore, C., Lepage, P., Ge, B., Bouchard, J.P., Mathieu, J., Melancon, S.B., Schalling, M., Lander, E.S., Morgan, K., Hudson, T.J. and Richter, A.
TITLE ARSACS, a spastic ataxia common in northeastern Quebec, is caused by mutations in a new gene encoding an 11.5-kb ORF
JOURNAL Nat. Genet. 24 (2), 120-125 (2000)
MEDLINE 20120709
REFERENCE 2 (bases 1 to 11493)
AUTHORS Engert, J.C., Berube, P., Dore, C., Lepage, P., Ge, B., Hudson, T.J. and Richter, A.
TITLE Direct Submission
JOURNAL Submitted (08-OCT-1999) Genome Centre, Montreal General Hospital, 1650 Cedar Ave., Montreal, QC H3G 1A4, Canada
FEATURES
source
1. 11493
/organism="Mus musculus"
/db_xref="taxon:10090"
<1. >11493
/product="sacsin"
1. 11493
/note="molecular chaperone"
/note="start=1"
/product="sacsin"
/protein_id="AAF31263.1"
/db_xref="GI:6907044"
/translation="MTFWGRELIVQWYFSEDRKRPSTLWLMKVMKNIYTHFSEDL TLDEMPILPTLLNDQRCVELIRLRISVYLIDETFAQIPEPLADIVOKRGIVL KRLDSIQHPVKKYIHSPLSAIIQIMKRIQLQKCNQIASILPHKRAKFLAST TPTSEKRIIOELTFKRIINSSDGISSYTKLGGCVLDTAKIPTDLRSVSD NSDEATIRLANMKTEKTKTSCLKFVLIDNNAFYTOEVLQMLMILENLSKNE NSNVLDWMLPDLKFHMSOGHVAAGDLPDIEVLDLFYNEEACFPPTIFTSPTIL HSLROIGKNESSEKEDVAVAKIEALQVSSCONQDVLMMKAKTLLVLNKNOTLL OSSGKMLAKTKWPAKCRPPNPGSLVWVGDCNLCAPDMCDAAHAYVGSSTP LYESVAVNIQALSTFTKPTINAVLKHFTVVDWTSKTFSEBDIYOFDHILLEITGF MHDHLSGKDSFRALKFPWWTGKNCPGLAOVAKFTDLDLQPYLVNPKTMAKPHQ

LPRAGSIEELTSDHISNVIQKYLKSDQELSEESKONLHLMNIMRLYSNQIPAS
PNPVPYIHSRNPSSKLMKPIHECCYCDIKVDLNDLDESVETLLVEDDIPMKRAE
WIKVPCILSTRNPENMGFSGOEPRLPIKNDLNEEYSPVSDILHEDDANNA
TCSGEMIDMRNMDIRENLDLPGMACHPALMSFNSEFSQDFNIRLDESLKRG
EVDKVGSELGNSVYHTIDPIIMREPMINDPNIHISKHINDRSPGJIKIMSK
OQRKLRKPNQEPKPIIDVFCQPLVAVPYSNGLFLPSFTQOEAIVSEVSTCY
NMADIYSLDEKSLCHRLIFQSVNSMYLYKLTIEFNLSADOTIIRKVCPSK
ALNAPYLSLDEAKLAKMKTCSSNNKLPDVKSSCIIOTIEEFHVRRIADLOSP
LFRGPDDPATFEEMAKSGOSKSPDELQKVCOTYTWLTCMDEMLKSLNSG
RLGIVGCAVYLHEHQBOKWTVPHGVEPCYPLIKIGLPHINGCAVYNSR
KEIMKTDITGRNNTTMRVYIKAYIQALSVRLDLAGSELIDYTYAYWPPDLYHD
DSVICKGFYEDIAHGKGLTRFVSDGSMWAMKVRFLDSDIQRKVGSAKFLIE
LYIKTKGSKNLCAYELPSSVKAAGEEACQIILENTESEKQFSEVFPNIOELA
ELRDPLMNVLENEKLEDFESGILRVTPCVCSLEGLVLPVSIHREGVAPLFDK
GREPVSTODYLNPIILIKIYOLGMAKDIDIMDMLHREASVAIKSHQAACLSKS
IILSIDEKIKIKDPRAKDPAKQYIPLPLTTRAGSLFKMGNSFPEMFAATD
ITYAEVQDVCLOPLINNSHSPFGCSVSLAVKEFLDLKKPPYDVIYNOLQYAK
SVDDGITLVQENTNACVYLYHEAVLQNMARATTEKLPCEPILVENAVYSEKVS
PHLNEAPAYLQLPNRYKNRRELSESVQSFVEDFALVLESIDERKQITTE
ENFOLCRITISEGIMSLIREKROEFCENYKILPDTMLLPLPASKLYNDCPWIKY
KDSVYRCHADIPREYAVKLGAIIPKRAKALERVASNICETALGETEGOKELTISR
ILNAVPEKEMILKEILLONADDAKATETICVPPROHPVDRIEDDKAPLOGLACVYN
NPFTEDDVRIQNLGKTRGEGNPKTGHTGIGFNSVYHTTQSPSTISNDLIGTDP
HARYAGATSVSPGRNFRDLADPRTQFSDVLDLYGNHFKLDNCTWFERPPLNAMA
OVSEISSVPSSDRMVONLIDKLSRSDGAELEMLNHEKISTICEIDKATGLVLYSVK
GKITDGDRLKRRQFASVDSVYKROLDIPOQYITTMDEDESGNLTWLYKRS
GFSSMEKVSYSIAHKNODITLPPRGVAACTTHYKKPHRAFCEPLSLEGLPEH
VNGHFDALDRNIMRDNGVGRSDMNSMTALIALAPYVELLITOLKRYRPGSPPT
LSYLONTPIHYVAKDTLKRFLSPPVARNLSDQDLYKVALYSICHEDMKRLLPVYA
PNIDGSDLSHVAITWINSTSKTRPFEDNLIDQDLQILKKNADYITRKRIAVNY
RLKHLLEIGFNLVYNCDETANLYHCLVADIPVSYTPADVRSFMTSPSPDNCHI
GKLPCLQOOTNLKLFHSLKLVQDCEFKDAEESFEVEGLPILITLDSVQIDGKRPK
PLTYHELIPSRKOLEMNTLYKSSVILNCQVAKYFSDSFDLSLSVLPREYKKN
CAKMDNFASSEMKNAMHETISESVSTDOEPRKAPFVIVYDIILKMDALLGTRFTY
STSQLVPEGCVLILPLSMHIAVFPNAGSDKRFHAKMGCIQOLANKICSDSALVP
LISCHTANDSPASILKAVHWOVSTFTEFKLEDEFPALLMTEFNCSHMSODDI
KILKSLPCYKSIISGRYMSIAKGTQCVLRSJPSAVEKMTOSSSAFSEBUEHLEL
YEVLGCVPDVDELYLHLPLKIENTSYAKLEHLYLKNRLASLESEPEIKOLEAL
LBSLIIITHANNRNKOAKHEVRYVVEPVEPKFIEKFEKFLKVOYIKRNOAF
MWSWVEFLNIGLITKALISOOLOLPAKELISVANTENMSEKFLQSTVYDILHIIQER
MDLSNPFLEKELSLIFLCEPBARPASYIRFHQOYQVNTPLIRFNQGVAPKRCQC
DVQILNMTISCPILPEKATPLSIREDSGLAQEOLEOVULNMLNVLDQVKNVINC
RINCIETLADIEEMVTRAKVLRISYIEFLAEKREFEFOIGVAFVWDEGMKLEFEE
VYINLEYEADFPYIKLPLELGETFHOLEKHGTEDIIISTKOVEVLSRIFSSSEKQ
LDPNEMRTKRVVSGLEFSLONDSYVRSDLEARNALALYLPDSODKLVKSSILEVD
APHYKSRIOGNIGVOMLVDSOCYIGKHGFHKLMLFPOKRLPRLSILIEODE
EPPKVCQFALCSLOGRDLILSSBPTTIGIRIKHENDNAFLANEKATIRCLARL
EGKAVSCFEKLDITTLKVGFPNTPHRSSTFAPLKRFGNAVILITIQHSDSKDINFL
ALAMTLKASATDNLISDTSYLIAMGNDIYRASEKLSAGVYDSESPKLDLPMGT
PIPAEIHVTLNDPMNVFPGEVGLVDAEGDIIYSGQPYTYTAIIOVEYERBAD
NPSFLKRIYOIDIGYSEKIVSGLDYKFSRDESSONDSAPPTPTSPTEFLFTGLR
SIPLSESGKESHSPSTKHSRPLKVNLAPELKEFLVSVQAMKLPSEKRIIIR
LYIKWHPDKNPNHDIANVEFKHLEIRLQKLEFVSVQAMKLSRSPSTASRPOS
DKYSRORTYSNQEATSHKSERODSKKCPSPSAGQYISQFEPYPTPKSGNPEVA
RKLRLQARANFSAARNDLHKNANENWCFQYISTKIALIALADYAVRGSKDQVKKPAL
AQKIEEYSOOLEGLTNDVHTLTAAYGWSLKTTRYPDLLPPOIIPNDRFSEVAMRWEC
TACIIIKLENFIOQKV"

BASE COUNT 3599 a 2281 c 2387 g 3226 t
ORIGIN
Query Match 3.8%; Score 53; DB 10; Length 11493;
Best Local Similarity 100.0%; Pred. No. 8e-17;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 298 tacacaacaccagcatttacagaagatgtagaggaattcaagaatcttg 350
|||||
Db 5521 TACAACACACGACCATTTACAGAAGATGATGTTAGAGGAATTCGATCTTGG 5573
|||||
RESULT 7
AC022025/c
LOCUS AC022025 189760 bp DNA linear PRI 15-DEC-2001

DEFINITION Homo sapiens chromosome 10 clone RP11-489D8, complete sequence.
AC020225
VERSION AC020225.6 GI:17861011
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 189760)
REFERENCE Smith,D.R.
AUTHORS Smith,D.R.
TITLE Genome Therapeutics Corporation Sequencing Center: Human Genome
JOURNAL Sequence Data
REFERENCE 2 (bases 1 to 189760)
AUTHORS Smith,D.R.
TITLE Direct Submission
JOURNAL Submitted (25-JAN-2000) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02453, USA
3 (bases 1 to 189760)
REFERENCE Smith,D.R.
AUTHORS Smith,D.R.
TITLE Direct Submission
JOURNAL Submitted (15-DEC-2001) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02453, USA
On Dec 15, 2001 this sequence version replaced gi:12957675.
COMMENT location/Qualifiers
FEATURES
1. 189760
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-489D8"
/clone.lib="RP11-11"
BASE COUNT 61011 a 36277 c 33262 g 59210 t
ORIGIN
Query Match 1.6%; Score 23; DB 9; Length 189760;
Best Local Similarity 100.0%; Pred. No. 0.54;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1255 aataacagtttaatgacagcatt 1277
|||||
Db 184482 AATAACAGTTTAAATGACGACATT 184460
RESULT 8
LOCUS AL356967 163869 bp DNA linear PRI 09-FEB-2001
DEFINITION Human DNA sequence from clone RP11-427E4 on chromosome 6, complete
sequence.
ACCESSION AL356967
VERSION AL356967
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 163869)
REFERENCE Tracy,A.
AUTHORS Tracy,A.
TITLE Direct Submission
JOURNAL Submitted (08-FEB-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
On Feb 9, 2001 this sequence version replaced gi:11691497.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
EM: EMBL; SW: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information
on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr6
RP11-427E4 is from the library RP11-11 constructed by the group
of Pletier de Jong. For further details see
http://www.chori.org/dacpac/home.htm
VECTOR: pBACE3.6
This sequence was finished as follows unless otherwise noted: all
regions were either double stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. This sequence is the
entire insert of clone RP11-427E4 The true right end of clone
RP11-685G11 is at 54518 in this sequence.
location/Qualifiers
1. 163869
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone="RP11-427E4"
/clone.lib="RP11-11.2"
52. .95
/note="5S repeat: matches 19. .65 of consensus"
203. .250
/note="L2 repeat: matches 2463. .2510 of consensus"
340. .461
/note="MIR repeat: matches 65. .192 of consensus"
467. .953
/note="MIR2C8 repeat: matches 1. .501 of consensus"
2103. .2466
/note="L2 repeat: matches 2127. .2487 of consensus"
2658. .2734
/note="L2 repeat: matches 1772. .1842 of consensus"
3382. .3621
/note="L1PA4 repeat: matches 5907. .6146 of consensus"
3622. .3815
/note="97 copies 2 mer aa 60% conserved"
7089. .7132
/note="22 copies 2 mer ta 100% conserved"
7676. .7778
/note="MIR repeat: matches 34. .138 of consensus"
8029. .8907
/note="L1MA8 repeat: matches 288. .1284 of consensus"
8909. .8974
/note="L1MA8 repeat: matches 4697. .4760 of consensus"
8975. .9276
/note="Alusx repeat: matches 1. .302 of consensus"
9277. .10794
/note="L1MA8 repeat: matches 4760. .6264 of consensus"
10806. .11071
/note="L1MD repeat: matches 1278. .1535 of consensus"
11137. .11260
/note="L2 repeat: matches 2579. .2710 of consensus"
12152. .12571
/note="L2 repeat: matches 72. .517 of consensus"
13044. .13865
/note="MER21B repeat: matches 1. .789 of consensus"
15518. .15664
/note="MIR repeat: matches 57. .213 of consensus"
15761. .16153
/note="MIR repeat: matches 1. .401 of consensus"
16284. .16415
/note="Alu/Alu/PRAM repeat: matches 176. .312 of consensus"
17048. .17352
/note="Alu repeat: matches 1. .305 of consensus"
17732. .18043
/note="Alu repeat: matches 1. .298 of consensus"
18174. .18393
/note="L1PA7 repeat: matches 5924. .6145 of consensus"

```

repeat_region 19046..19188
/note="MLT1J repeat: matches 129. .273 of consensus"
repeat_region 22980..23144
/note="L2 repeat: matches 2570. .2707 of consensus"
repeat_region 25307..25360
/note="27 copies 2 mer tt 72% conserved"
repeat_region 25620..25714
/note="L1M4 repeat: matches 6183. .6250 of consensus"
repeat_region 25715..25996
/note="AluJo repeat: matches 1. .282 of consensus"
repeat_region 25997..26027
/note="L1M8 repeat: matches 6250. .6285 of consensus"
repeat_region 26066..26259
/note="L1MC5 repeat: matches 7336. .7549 of consensus"
repeat_region 26522..26629
/note="MIR repeat: matches 76. .189 of consensus"
repeat_region 26767..27106
/note="MSTD repeat: matches 1. .394 of consensus"
repeat_region 27238..27847
/note="L1PA4 repeat: matches 5210. .5819 of consensus"
repeat_region 27848..28190
/note="L1PA4 repeat: matches 5800. .6142 of consensus"
repeat_region 29252..29372
/note="MIR repeat: matches 76. .200 of consensus"
repeat_region 30389..30552
/note="MER63A repeat: matches 3. .174 of consensus"
repeat_region 31093..31222
/note="MIR repeat: matches 13. .146 of consensus"
repeat_region 31943..31984
/note="14 copies 3 mer tta 78% conserved"
repeat_region 32356..32545
/note="L1R33 repeat: matches 329. .521 of consensus"
repeat_region 32696..32897
/note="MER33 repeat: matches -10. .65 of consensus"
repeat_region 32994..32994
/note="L1R33 repeat: matches 90. .141 of consensus"
repeat_region 33597..33687
/note="MIR repeat: matches 56. .148 of consensus"
repeat_region 33726..33807
/note="L2 repeat: matches 2624. .2708 of consensus"
repeat_region 33843..33923
/note="27 copies 3 mer ata 65% conserved"
repeat_region 34008..34082
/note="MIR repeat: matches 66. .141 of consensus"
repeat_region 35184..35463
/note="AluJo repeat: matches 1. .288 of consensus"
repeat_region 36308..36729
/note="L2 repeat: matches 2229. .2695 of consensus"
repeat_region 37243..37308
/note="33 copies 2 mer aa 66% conserved"
repeat_region 38342..38636
/note="L1M4 repeat: matches 3889. .4188 of consensus"
repeat_region 38640..38678
/note="L1R40b repeat: matches 422. .462 of consensus"
repeat_region 38679..39027
/note="L1PA7 repeat: matches 5787. .6137 of consensus"
repeat_region 39028..39154
/note="L1R40b repeat: matches 308. .422 of consensus"
repeat_region 39159..39571
/note="L1M4 repeat: matches 2823. .3283 of consensus"
repeat_region 39626..39665
/note="20 copies 2 mer at 77% conserved"
repeat_region 40095..40393
/note="AluSc repeat: matches 5. .301 of consensus"
repeat_region 40934..41565
/note="L1MC6 repeat: matches 1203. .1921 of consensus"
repeat_region 41925..42147
/note="L2 repeat: matches 2466. .2745 of consensus"
repeat_region 42969..43268
/note="AluJb repeat: matches 1. .293 of consensus"
repeat_region 43667..43708
/note="21 copies 2 mer tg 100% conserved"
repeat_region 44902..45177

```

```

repeat_region /note="AluJo repeat: matches 1. .298 of consensus"
46349..46454
/note="53 copies 2 mer at 62% conserved"
repeat_region 47391..47625
/note="MIR repeat: matches 13. .262 of consensus"
repeat_region 47615..47663
/note="MIR repeat: matches 210. .252 of consensus"
repeat_region 47637..47678
/note="L2 repeat: matches 2708. .2750 of consensus"
repeat_region 50360..50596
/note="MIR repeat: matches 8. .261 of consensus"
repeat_region 51156..51205
/note="25 copies 2 mer ta 72% conserved"
repeat_region 51347..51733
/note="MLT1D repeat: matches 69. .500 of consensus"
repeat_region 51734..52036
/note="AluJb repeat: matches 1. .286 of consensus"
repeat_region 52037..52110
/note="MLT1D repeat: matches 2. .69 of consensus"
repeat_region 52237..52752
/note="L2 repeat: matches 1600. .2152 of consensus"
repeat_region 52768..53179
/note="L1R16A repeat: matches 25. .442 of consensus"
repeat_region 53191..53668
/note="L2 repeat: matches 2151. .2750 of consensus"
repeat_region 53672..53886
/note="MIR repeat: matches 19. .243 of consensus"

```

```

Query Match 1.6% Score 22; DB 9; Length 163869;
Best Local Similarity 100.0%; Pred. No. 1.9;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1129 ttctgttttgcctctctt 1150
Db 94539 ttctgttttgcctcttctt 94560

```

```

RESULT 9
AC013913 2296 bp DNA linear HTG 16-NOV-1999
LOCUS AC013913.C
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
pieces.
ACCESSION AC013913
VERSION AC013913.1 GI:6437422
KEYWORDS HTG; HTGS; PHASE2.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 2296)
REFERENCE Adams M. and Venter, J.C.
AUTHORS Direct Submission
JOURNAL Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA

```

COMMENT This sequence was identified as CDW:10213920 by the submitter.
For further information on this sequence e-mail to fly@celera.com.

* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES
source 1..2296
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"

BASE COUNT 891 a 384 c 381 g 640 t
ORIGIN

Query Match 1.5% Score 21; DB 2; Length 2296;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1321 ggtattccctggtctgac 1341
 |||||||
 Db 1248 GGTATTCCCTGCTTGATC 1228

RESULT 10
 AL136322
 LOCUS
 DEFINITION Human DNA sequence from clone RP11-75C23 on chromosome 1q31.2-32.1,
 complete sequence.
 ACCESSION AL136322
 VERSION AL136322.24 GI:18072465
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 55954)
 Smith, M.
 Direct Submission

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 Submitted (04-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 humquery@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
 On Jan 6, 2002 this sequence version replaced g1:17902846.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >= 30);
 an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid subclone or more than one M13 subclone; and the
 assembly was confirmed by restriction digest. The following
 abbreviations are used to associate primary accession numbers given
 in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP
 database can be found at
 http://www.sanger.ac.uk/Projects/C.elegans/wormpep This sequence
 was generated from part of bacterial clone contigs of human
 chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
 Group. Further information can be found at
 http://www.sanger.ac.uk/HGP/Chr1
 RP11-75C23 is from the library RPCI-11.1 constructed by the group
 of Pieter de Jong. For further details see
 http://www.chori.org/bacpac/home.htm
 VECTOR: pBACE3.6

FEATURES
 source
 1..55954
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="1"
 /map="q31.2-32.1"
 /clone="RP11-75C23"
 /clone_1lb="RPCI-11.1"
 BASE COUNT 17559 a 10247 c 10324 g 17824 t
 ORIGIN

Query Match 1.5%; Score 21; DB 9; Length 55954;
 Best Local Similarity 100.0%; Pred. No. 7.7;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1332 ggttcgacacacattatca 1352
 |||||||

Db 892 GGTTCGATCCACATTATCA 912

RESULT 11
 AC091528.1/c
 WPCOMMENT
 Sequence split into 5 fragments LOCUS AC091528 Accession AC091528

Fragment Name	Begin	End
AC091528.0	1	110000
AC091528.1	100001	210000
AC091528.2	200001	310000
AC091528.3	300001	410000
AC091528.4	400001	431537

Continuation (2 of 5) of AC091528 from base 100001 (AC091528 Homo sapiens chromosome

Query Match 1.5%; Score 21; DB 2; Length 110000;
 Best Local Similarity 100.0%; Pred. No. 6.9;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 227 ctgacagcatccagttgata 247
 |||||||
 Db 21772 CTGACAGCATCCAGTTGATA 21752

RESULT 12
 AL445469/c
 LOCUS
 DEFINITION Human DNA sequence from clone RP11-254N18 on chromosome 1, complete
 sequence.
 ACCESSION AL445469
 VERSION AL445469.4 GI:11991422
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 129504)
 Baggaley, C.
 Direct Submission

COMMENT
 Submitted (23-DEC-2000) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
 requests: clonerequests@sanger.ac.uk
 On Dec 24, 2000 this sequence version replaced g1:11879926.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 This sequence has been finished according to sequence map criteria
 as follows. An attempt is made to resolve all sequencing problems,
 such as compressions and repeats, but not necessarily within known
 annotated repeat sequence elements. Where the sequence is
 ambiguous, there is an annotation using the 'unseq' feature key.
 The following abbreviations are used to associate primary accession
 numbers given in the feature table with their source databases:
 Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information
 on the WORMPEP database can be found at
 http://www.sanger.ac.uk/Projects/C.elegans/wormpep This sequence
 was generated from part of bacterial clone contigs of human
 chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
 Group. Further information can be found at
 http://www.sanger.ac.uk/HGP/Chr1
 RP11-254N18 is from the library RPCI-11.1 constructed by the group
 of Pieter de Jong. For further details see
 http://www.chori.org/bacpac/home.htm
 VECTOR: pBACE3.6

IMPORTANT: This sequence is not the entire insert of clone
 RP11-254N18 it may be shorter because we sequence overlapping
 sections only once, except for a 100 base overlap.
 The true right end of clone RP11-254N18 is at 129504 in this
 sequence. The true right end of clone RP5-1180C10 is at 100 in this

sequence. The true right end of clone RP11-118D10 is at 95065 in this sequence.

FEATURES
Location/Qualifiers
1. 129504
source

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-254N18"
/clone_1ib="RPCT-11.1"
42809..42924
misc_feature

BASE COUNT 44541 a 24211 c 23194 g 37558 t
(AL359938). Assembly confirmed by restriction digest."

ORIGIN

Query Match 1.5%; Score 21; DB 9; Length 129504;
Best Local Similarity 100.0%; Pred. No. 6.8;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1177 aatgcacacttgactgcat 1197
|||||
Db 107258 AATGCCACTTTCACCTGCAAT 107238

RESULT 13
AC106679/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AC106679 157913 bp DNA linear HNG 12-JAN-2002
Rattus norvegicus clone CH230-106A6, *** SEQUENCING IN PROGRESS
*** 57 unordered pieces.

AC106679
HTG: HTGS PHASE1.
Norway rat.
Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
AUTHORS

1 (bases 1 to 157913)
Muzny,D.M., Adams,C., Adio-Oduola,B., All-rosman,F.R., Allen,C.,
Alshrooks,S.L., Amartunge,H.C., Are,J.R., Banks,T., Barberia,J.,
Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowle,S., Brieve,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burke,P., Burkett,C., Burrell,K.L., Byrd,N.C., Cartron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhury,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Doultwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garra,N., Gill,R., Gorrell,J.H., Guevara,W., Gunartine,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hognes,M., Holloway,C.,
Hollins,B., Homsl,F., Howard,S., Huber,J., Huliy,K., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,U., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Louisege,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhinney,E., McLeod,M.P., Meador,M.,
Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabat,K.,
Moran,M., Morris,S., Moser,W., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokweto,S.,
Ogun,M., Okunolu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojupokan,I., Rolfe,M.,
Ruiz,S., Savary,G., Scherer,S., Scott,G., Shen,H., Shoehart,N.,
Sisson,I., Sodergren,E., Sonalle,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Taber,P., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Washington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 157913)

Submitted (12-JAN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
Project Information
Center project name: GLUO
Center clone name: CH230-106A6

Summary Statistics
Assembly program: Phrap; version 0.990329first call to
findphraplist

Consensus quality: 143594 bases at least Q40
Consensus quality: 150694 bases at least Q30
Consensus quality: 156627 bases at least Q20
Estimated insert size: 142942; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-tp estimation
Quality coverage: 2.8x in Q20 bases; sum-of-contigs estimation

NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).

NOTE: This is a 'working draft' sequence. It currently
consists of 57 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1
5729: contig of 5729 bp in length
5730 5829: gap of unknown length
5830 12058: contig of 6229 bp in length
12059 12158: gap of unknown length
12159 17377: contig of 5219 bp in length
17378 17477: gap of unknown length
17477 23186: contig of 5709 bp in length
23186 23286: gap of unknown length
23286 28405: contig of 5119 bp in length
28405 33071: contig of 4566 bp in length
33071 33171: gap of unknown length
33171 37962: contig of 4791 bp in length
37962 38062: gap of unknown length
38062 42336: contig of 4274 bp in length
42336 42436: gap of unknown length
42436 47541: contig of 5105 bp in length
47541 47641: gap of unknown length
47641 51053: contig of 3412 bp in length
51053 51153: gap of unknown length
51153 54397: contig of 3244 bp in length
54397 54497: gap of unknown length
54497 57140: contig of 2643 bp in length
57140 57240: gap of unknown length
57240 60305: contig of 3065 bp in length
60305 60405: gap of unknown length
60405 64988: contig of 4583 bp in length
64988 65089: gap of unknown length
65089 69178: contig of 4090 bp in length
69178 71834: gap of unknown length
71834 71934: contig of 2556 bp in length
71934: gap of unknown length


```

* 71935 74740: contig of 2806 bp in length
* 74741 74840: gap of unknown length
* 74841 77660: contig of 2820 bp in length
* 77661 77760: gap of unknown length
* 82112 82211: contig of 4351 bp in length
* 82212 82211: gap of unknown length
* 86065 86065: contig of 3854 bp in length
* 86066 86165: gap of unknown length
* 86166 88690: contig of 2525 bp in length
* 88691 88790: gap of unknown length
* 88791 91103: contig of 2313 bp in length
* 91104 91203: gap of unknown length
* 91204 94206: contig of 3003 bp in length
* 94207 94306: gap of unknown length
* 94307 95998: contig of 1692 bp in length
* 95999 96098: gap of unknown length
* 96099 99075: contig of 2977 bp in length
* 99076 99175: gap of unknown length
* 99176 102430: contig of 3255 bp in length
* 102431 102530: gap of unknown length
* 102531 104203: contig of 1673 bp in length
* 104204 104303: gap of unknown length
* 104304 106512: contig of 2209 bp in length
* 106513 106612: gap of unknown length
* 106613 109297: contig of 2685 bp in length
* 109298 109397: gap of unknown length
* 109398 113020: contig of 3623 bp in length
* 113021 113120: gap of unknown length
* 113121 115477: contig of 2357 bp in length
* 115478 115577: gap of unknown length
* 115578 117671: contig of 2094 bp in length
* 117672 117771: gap of unknown length
* 117772 119975: contig of 2204 bp in length
* 119976 120075: gap of unknown length
* 120076 121881: contig of 1806 bp in length
* 121882 121981: gap of unknown length
* 121982 124296: contig of 2315 bp in length
* 124297 124396: gap of unknown length
* 124397 125748: contig of 1352 bp in length
* 125749 125848: gap of unknown length
* 125849 127673: contig of 1825 bp in length
* 127674 127773: gap of unknown length
* 127774 129419: contig of 1646 bp in length
* 129420 129519: gap of unknown length
* 130855 130855: contig of 1336 bp in length
* 130856 130955: gap of unknown length
* 130956 132588: contig of 1633 bp in length
* 132589 132688: gap of unknown length
* 132689 134332: contig of 1644 bp in length
* 134333 134432: gap of unknown length
* 134433 136267: contig of 1835 bp in length
* 136268 136367: gap of unknown length
* 136368 137731: contig of 1364 bp in length
* 137732 137831: gap of unknown length
* 137832 139895: contig of 2064 bp in length
* 139896 139995: gap of unknown length
* 139996 141525: contig of 1530 bp in length
* 141526 141625: gap of unknown length
* 141626 143293: contig of 1668 bp in length
* 143294 143393: gap of unknown length
* 143394 144551: contig of 1158 bp in length
* 144552 144651: gap of unknown length
* 144652 146085: contig of 1434 bp in length
* 146086 146185: gap of unknown length
* 146186 147247: contig of 1062 bp in length
* 147248 147347: gap of unknown length
* 147348 148580: contig of 1233 bp in length
* 148581 148680: gap of unknown length
* 148681 149786: contig of 1106 bp in length
* 149787 149886: gap of unknown length
* 149887 150935: contig of 1049 bp in length
* 150936 151035: gap of unknown length
* 151036 152387: contig of 1352 bp in length

```

```

* 152388 152487: gap of unknown length
* 152488 153785: contig of 1298 bp in length
* 153786 153885: gap of unknown length
* 153886 155369: contig of 1484 bp in length
* 155370 155469: gap of unknown length
* 155470 156335: contig of 1166 bp in length
* 156336 156735: gap of unknown length
* 156736 157913: contig of 1178 bp in length.

Query Match      1.5%: Score 21; DB 2; Length 157913;
Best Local Similarity 100.0%; Pred. No. 6.5;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1018 aaagatctaaagtgtcata 1038
Db 80552 AAAGATCTAAAGTGTCATA 80532

RESULT 14
AC084374/c
LOCUS
DEFINITION
HOMO sapiens chromosome 12q clone R11-13403, WORKING DRAFT
AC084374
AC084374
AC084374.20 GI:17105262
HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
KEYWORDS
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 171926)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-ozman,F.R., Allen,C.,
Alshrocks,S.L., Amaralunge,H.C., Are,J.R., Banks,T., Barbata,J.,
Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouk,T.,
Bowle,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Cartron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,K.,
Hollins,B., Homai,F., Howard,S., Huber,J., Huiyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W.,
Louisged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapa,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhney,E., Mcleod,M.P., Meador,M.,
Mel,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabac,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokikeko,S.,
Ogulu,M., Okwunou,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Oulles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,L.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoohari,N.,
Sisson,I., Sodergren,E., Sonake,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Swalek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Tellford,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Verra,V., Villalón,D., Vinsón,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Watlington,S., Williams,G., Williamson,A., Wleczky,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 171926)
TITLE
JOURNAL
REFERENCE

```

AUTHORS
TITLE
JOURNAL

Worley, K.C.
Direct Submission
Submitted (27-OCT-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 27, 2001 this sequence version replaced g1:16874816.

COMMENT

Center: Baylor College of Medicine
Genome Center
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

Project Information
Center project name: HCCJE
Center clone name: RP11-13403

Summary Statistics
Sequencing vector: Plasmid: M77789

Chemistry: Dye-terminator Big Dye: 99% of reads
Assembly program: Phrap; version 0.990329

Consensus quality: 173766 bases at least Q40

Consensus quality: 175002 bases at least Q20

Estimated insert size: 172132; sum-of-coverage estimation

Quality coverage: 0x in Q20 bases; agarose-gel estimation

Quality coverage: 6.4x in Q20 bases; sum-of-coverage estimation

* NOTE: Estimated insert size may differ from sequence length

* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently

* consists of 1 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

1 171926: contig of 171926 bp in length.

Location/Qualifiers

1. 171926

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="12q"

/clone="RP11-13403"

/size="31150 c 30825 g 52933 t

BASE COUNT 57018 a 31150 c 30825 g 52933 t

FEATURES
source

Query Match

Best Local Similarity 100.0%; Score 21; DB 2; Length 171926;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 227 ctgacgagcatcagtgata 247
|||||

Db 18991 CTGACGACATCCAGTTGATA 18971

RESULT 15

AC096964 176912 bp DNA linear HTG 20-DEC-2001

LOCUS Rattus norvegicus clone CH230-202E24, *** SEQUENCING IN PROGRESS

DEFINITION *** 67 unordered pieces.

AC096964 176912 bp DNA linear HTG 20-DEC-2001

AC096964 176912 bp DNA linear HTG 20-DEC-2001

AC096964 176912 bp DNA linear HTG 20-DEC-2001

AC096964 176912 bp DNA linear HTG 20-DEC-2001

AC096964 176912 bp DNA linear HTG 20-DEC-2001

AC096964 176912 bp DNA linear HTG 20-DEC-2001

AC096964 176912 bp DNA linear HTG 20-DEC-2001

AC096964 176912 bp DNA linear HTG 20-DEC-2001

AC096964 176912 bp DNA linear HTG 20-DEC-2001

AC096964 176912 bp DNA linear HTG 20-DEC-2001

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

Project Information

Center project name: GHRN

Center clone name: CH230-202E24

Summary Statistics

Assembly program: Phrap; version 0.990329first call to

findhaplolist

Consensus quality: 151663 bases at least Q40

Consensus quality: 160070 bases at least Q40

Consensus quality: 166999 bases at least Q20

Estimated insert size: 152238; sum-of-coverage estimation

Quality coverage: 0x in Q20 bases; agarose-gel estimation

Quality coverage: 2.3x in Q20 bases; sum-of-coverage estimation

* NOTE: Estimated insert size may differ from sequence length

* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently

* consists of 67 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buha, C.,
Carter, M., Cavazos, S.R., Chacko, K.L., Byrd, N.C., Caron, T.F.,
Chen, Z., Chowdhury, I., Christopoulos, C., Cleveland, C.D., Cox, C.,
Coyne, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C.,
Degen, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H.,
Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C.,
Eibach, C., Escotto, M., Falls, T., Ferraruto, D., Flagg, N., Ford, J.,
Foster, P., Frantz, P., Gabisi, A., Geo, J., Garcia, A., Garner, T.,
Garza, N., Gill, R., Gottlieb, D.H., Guevara, M., Gunaratne, P., Hale, S.,
Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A.,
Hernandez, O., Hernandez, O., Hodson, A., Hogue, M., Holloway, C.,
Hollins, B., Homsl, F., Howard, S., Huber, J., Hulik, S., Hume, J.,
Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S.,
Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J.,
Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C.,
Lewis, L., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W.,
Lounsbury, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R.,
Ma, J., Maheshwari, M., Mapa, P., Martin, R., Martindale, A.,
Martinez, E., Massey, E., Mawhinney, E., McLeod, M.P., Meador, M.,
Mei, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabab, K.,
Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newsham, N.,
Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokwenkwo, S.,
Ogih, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B.,
Peery, J., Perez, L., Peters, R., Pickens, R., Primus, E., Pu, L.L.,
Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M.,
Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shooshari, N.,
Stinson, I., Sodergren, E., Sonalke, T., Sparks, A., Stanley, H.,
Stone, H., Sutton, A., Svatek, A., Taber, P., Tamerisa, A., Tamerisa, K.,
Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N.,
Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R.,
Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,
Washington, S., Williams, G., Williamson, A., Wleciak, R., Wooden, S.,
Worley, K., Wu, C., Wu, Y.F., Zhou, J., Zorilla, S., Nelson, D.,
Weinstock, G., and Gibbs, R.

Direct Submission

Unpublished

2 (bases 1 to 176912)

Worley, K.C.

Submitted (04-OCT-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

On Dec 20, 2001 this sequence version replaced g1:17064280.

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

Project Information

Center project name: GHRN

Center clone name: CH230-202E24

Summary Statistics

Assembly program: Phrap; version 0.990329first call to

findhaplolist

Consensus quality: 151663 bases at least Q40

Consensus quality: 160070 bases at least Q40

Consensus quality: 166999 bases at least Q20

Estimated insert size: 152238; sum-of-coverage estimation

Quality coverage: 0x in Q20 bases; agarose-gel estimation

Quality coverage: 2.3x in Q20 bases; sum-of-coverage estimation

* NOTE: Estimated insert size may differ from sequence length

* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently

* consists of 67 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

1 7294: contig of 7294 bp in length
7295 7394: gap of unknown length
12823 12833: contig of 5429 bp in length
12824 12923: gap of unknown length
12924 17505: contig of 4582 bp in length
17506 17605: gap of unknown length
17606 24532: contig of 6927 bp in length
24533 24632: gap of unknown length
24633 29839: contig of 5207 bp in length
29840 29939: gap of unknown length
29940 35829: contig of 5890 bp in length
35830 35930: gap of unknown length
35931 39896: contig of 3967 bp in length
39897 44844: gap of unknown length
44845 44944: contig of 4848 bp in length
44945 50921: contig of 5977 bp in length
50922 51021: gap of unknown length
51022 54518: contig of 3497 bp in length
54519 54618: gap of unknown length
54619 59782: contig of 5164 bp in length
59783 59882: gap of unknown length
59883 63190: contig of 3307 bp in length
63190 63289: gap of unknown length
63290 67064: contig of 3775 bp in length
67065 70204: gap of unknown length
70205 70304: contig of 3040 bp in length
70305 73659: contig of 3355 bp in length
73660 73759: gap of unknown length
73760 76600: contig of 2841 bp in length
76601 76700: gap of unknown length
76701 81449: contig of 4749 bp in length
81450 81549: gap of unknown length
81550 84603: contig of 3054 bp in length
84604 84703: gap of unknown length
84704 87038: contig of 2335 bp in length
87039 87138: gap of unknown length
87139 88733: contig of 1595 bp in length
88734 88833: gap of unknown length
88834 92170: contig of 3337 bp in length
92171 92270: gap of unknown length
92271 94942: contig of 2672 bp in length
94943 95042: gap of unknown length
95043 97980: contig of 2938 bp in length
97981 98080: gap of unknown length
98081 101026: contig of 2946 bp in length
101027 101126: gap of unknown length
101127 103481: contig of 2355 bp in length
103482 103581: gap of unknown length
103582 105942: contig of 2361 bp in length
105943 106042: gap of unknown length
106043 108765: contig of 2723 bp in length
108766 108865: gap of unknown length
108866 111347: contig of 2482 bp in length
111348 111447: gap of unknown length
111449 114803: contig of 3356 bp in length
114804 114903: gap of unknown length
114904 116772: contig of 1869 bp in length
116773 116872: gap of unknown length
116874 118340: contig of 1468 bp in length
118341 118440: gap of unknown length
118441 120587: contig of 2147 bp in length
120588 120687: gap of unknown length
120689 122289: contig of 1602 bp in length
122290 122389: gap of unknown length
122390 124086: contig of 1697 bp in length
124087 124186: gap of unknown length
124187 126276: contig of 2090 bp in length
126277 126376: gap of unknown length
126377 127942: contig of 1566 bp in length
127943 128042: gap of unknown length
128043 129220: contig of 1178 bp in length

129221 129320: gap of unknown length
129321 130478: contig of 1158 bp in length
130479 130578: gap of unknown length
130579 131805: contig of 1227 bp in length
131806 131905: gap of unknown length
131906 133279: contig of 1374 bp in length
133280 133379: gap of unknown length
133380 135075: contig of 1696 bp in length
135076 135175: gap of unknown length
135176 137317: contig of 2142 bp in length
137318 137417: gap of unknown length
137418 138750: contig of 1333 bp in length
138751 138850: gap of unknown length
138851 141412: contig of 2562 bp in length
141413 141512: gap of unknown length
141513 142664: contig of 1152 bp in length
142665 142764: gap of unknown length
142765 144121: contig of 1357 bp in length
144122 144221: gap of unknown length
144222 146975: contig of 2754 bp in length
146976 147075: gap of unknown length
147076 148195: contig of 1120 bp in length
148196 148295: gap of unknown length
148296 149533: contig of 1238 bp in length
149534 149633: gap of unknown length
149634 151328: contig of 1695 bp in length
151329 151428: gap of unknown length
151429 153745: contig of 2317 bp in length
153746 153845: gap of unknown length
153846 155423: contig of 1578 bp in length
155424 155523: gap of unknown length
155524 157366: contig of 1843 bp in length
157367 157466: gap of unknown length
157467 159157: contig of 1691 bp in length
159158 159257: gap of unknown length
159258 160289: contig of 1032 bp in length
160290 160389: gap of unknown length
160390 161589: contig of 1200 bp in length
161590 161689: gap of unknown length

Query Match 1.5%; Score 21; DB 2; Length 176912;
Best Local Similarity 100.0%; Pred.No. 6.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0;
Gaps 0;
OY 24 cccaagcgacacaagcctt 44
|||||
Db 82779 CCAAGGACACAAAGCCTT 82799

Search completed: May 22, 2002, 08:46:53
Job time: 11703 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 22, 2002, 06:36:01 ; Search time 3530.57 Seconds
(without alignments)
8304.064 Million cell updates/sec

Title: US-09-693-205-7_COPY_5300_6700

Perfect score: 1401
Sequence: 1 gtacgacgaaactagagc.....aaagacacttaagaagt 1401

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*
1: gb.ba:*
2: gb.htg:*
3: gb.in:*
4: gb.om:*
5: gb.ov:*
6: gb.pat:*
7: gb.ph:*
8: gb.pl:*
9: gb.pr:*
10: gb.ro:*
11: gb.stg:*
12: gb.sy:*
13: gb.un:*
14: gb.vl:*
15: em.ba:*
16: em.fun:*
17: em.hum:*
18: em.in:*
19: em.mu:*
20: em.om:*
21: em.ot:*
22: em.ov:*
23: em.pat:*
24: em.ph:*
25: em.pl:*
26: em.ro:*
27: em.sts:*
28: em.un:*
29: em.vl:*
30: em.htg.hum:*
31: em.htg.inv:*
32: em.htg.other:*
33: em.htgo.inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
------------	-------	-------	-------	--------	----	----	-------------

1	1390	99.2	12793	6	AX119931	AX119931 Sequence
2	1390	99.2	12793	9	AF193556	AF193556 Homo sapi
3	1390	99.2	92693	9	AL157766	AL157766 Human DNA
4	1364.6	97.4	99819	2	AC079761	AC079761 Homo sapi
5	1174	83.8	11492	6	AX119933	AX119933 Sequence
6	1174	83.8	11493	10	AF193557	AF193557 Mus muscu
7	126.6	9.0	78844	8	AB006708	AB006708 Arabidops
8	115.2	8.2	165299	2	OSJN00002	OSJN00002
9	58	4.1	171410	2	AC069017	AC069017 Oryza sat
10	54.4	3.9	318221	2	PFMA01363	PFMA01363 Mus muscu
11	51.8	3.7	1141	6	AX083744	AX083744 Plasmodiu
12	45	3.2	2389	3	AF151112	AF151112 Dictyoste
13	45	3.2	3885	3	AF151733	AF151733 Dictyoste
14	45	3.2	3985	3	AF151111	AF151111 Dictyoste
15	44.6	3.2	160203	9	HS425C14	HS425C14 Human DNA s
16	44.4	3.2	114532	2	AC016547	AC016547 Homo sapi
17	44.4	3.2	129046	9	AC008835	AC008835 Homo sapi
18	44.4	3.2	143358	2	AC044843	AC044843 Homo sapi
19	44.4	3.2	161133	2	AC019295	AC019295 Homo sapi
20	44.4	3.2	163690	2	AC083781	AC083781 Homo sapi
21	44	3.1	282928	2	AC091244	AC091244 Rattus no
22	43.4	3.1	187413	9	AC007567	AC007567 Homo sapi
23	43.2	3.1	7218	6	166494	166494 Sequence 14
24	43.2	3.1	83650	8	AB023041	AB023041 Arabidops
25	43.2	3.1	187193	2	AC105657	AC105657 Rattus no
26	43	3.1	231450	1	AL596163	AL596163 Listeria
27	42.6	3.0	169019	2	AC078952	AC078952 Homo sapi
28	42.6	3.0	243675	10	AC079869	AC079869 Mus muscu
29	42	3.0	199890	2	AC095674	AC095674 Rattus no
30	41.6	3.0	102406	2	AC103075	AC103075 Rattus no
31	41.4	3.0	110000	2	AC011600_1	Continuation (2 of
32	40.8	2.9	35338	8	CAC49C4	AL033503 C.albican
33	40.8	2.9	167941	9	AC092638	AC092638 Homo sapi
34	40.8	2.9	234112	3	PFMAL4P2	AL033475 Plasmodiu
35	40.6	2.9	7990	6	AX281265	AX281265 Sequence
36	40.6	2.9	7990	6	AX345060	AX345060 Sequence
37	40.6	2.9	28125	3	U40426	U40426 Caenorhabdl
38	40.4	2.9	50886	9	AL450107	AL450107 Human DNA
39	40.4	2.9	70326	2	AC068957	AC068957 Homo sapi
40	40.4	2.9	161714	2	AC092087	AC092087 Canis fam
41	40.4	2.9	170928	9	AC034148	AC034148 Homo sapi
42	40.4	2.9	190937	9	AP000957	AP000957 Homo sapi
43	40.4	2.9	340000	9	AP001680	AP001680 Homo sapi
44	40.2	2.9	122241	2	AP004129	AP004129 Oryza sat
45	40	2.9	52706	2	AC101794	AC101794 Mus muscu

ALIGNMENTS

RESULT	1	LOCUS	AX119931	Sequence 1	12793 bp	DNA	Linear	PAT 11-MAY-2001
DEFINITION	AX119931	AX119931	Sequence 1 from Patent WO0129266.					
ACCESSION	AX119931	AX119931						
VERSION	AX119931.1	GI:14036678						
KEYWORDS								
SOURCE		human.						
ORGANISM		Homo sapiens						

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 12793)

AUTHORS Hudson,T.J., Engert,J. and Richter,A.

Query Match 99.2%; Score 1390; DB 6; Length 12793;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1401; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

```

QY 1 gtagcagtaaaactagagcagtcaccaagaacacaaagccttagaagaatgcattcc 60
DB 5300 GTTAGCAGTAAACCTAGGAGCAGTCCCAAGCGACACAAACCTTAGAATATGATCC 5359
QY 61 aatgctgttttcaaacacttgacagaaattgggcaagaataaattgaccagaga 120
DB 5360 AATGCTGTTTAAACACTTGGCAGCAAAATTTGGCAGAAAGAAATTAAGACAGAGA 5419
QY 121 attaagaacatcccttaatacatalcctctgaaagaatgltgaaagccttctca 180
DB 5420 ATTAAGAGCATCCTTAATGATATCCTCTGAAAGAAATGTTGAAGAGCTTCTTCA 5479
QY 181 aatgctgataagcaaaagcgacagaatctglttgltgltgacccatagacacatcca 240
DB 5480 AATGCTGATGATCAAAAGCGACAGAAATCTGTTTGTGTGATCCTTAGACAGCATCCA 5539
QY 241 gttgataagaatattgataagtgagcccatgcaagggcagcagacttgggtac 300
DB 5540 GTTGATAGAAATATTTGATGATAAGTGCGCCCATGTCAGAGGCGACACTTGTGTATC 5599
QY 301 aacaacacagccattcaagaagaatglttagagaatcagaatcttgaaagaagcag 360
DB 5600 AACACACAGCATTTACAGAGATGATGTTAGAGAAATTCAGAAATCTTGAAAAAGCAGC 5659
QY 361 aaagagggaatccctataaactagacagatgaaatagatcaattctgltacat 420
DB 5660 AAAGAGGGAATCTTATTAACCTGAGACAGATGAGAAATGAGAAATCTGTTATCAT 5719
QY 421 atcacagactgcccacttattcttgcaatgacatccctggtatttggctccat 480
DB 5720 ATCACGACTGCCCATCTTTATTTCTGGCAATGACATCCCTGTTATTTTGATCCCAT 5779
QY 481 gccacatagcacaggggccaacatcatcagtcgccgagcagatglttagaattgag 540
DB 5780 GCCAGATATGACACAGGCGGCACATCATATAGTCCGAGAGCATGTTTAGAGTTTGAT 5839
QY 541 gcagattttaagaacacagttctcagatgltcagatccttattctgggaaccatttaa 600
DB 5840 GCAGATTTTAAAGACAGATTTCTCAGATGTTCTGATCTTATCTGGCAACCCATTTTAA 5899
QY 601 ctgataattgacacaatglttcagattcctcttcgtaagtcaagaatggaaagttcg 660
DB 5900 CTGATATATGCAATGTTCAAGATTTCTCTTGATGACAGAAATGGCAAAAGTTTCG 5959
QY 661 gaaattctgltcttcacagatcagacagaatgltccagaatcttltgacaactgcg 720
DB 5960 GAAATTTTCGTCGTTCAGATCAGACAGAAAGTCCAGAAATCTTTTGCAAAAGTCCG 6019
QY 721 tcagatggggcagaactcttaagtcttcaatcacatgtaaaaaattctatttgtaa 780
DB 6020 TCAGATGGGCGCAAACTTTAAATGTTCTTAAATCACATGAAAAAATTTCTATTGTGAA 6079
QY 781 atagataagagtagtgcagctcaaatgtagctgattcagtaaaggccaatcacagat 840
DB 6080 ATAGATAAGAGTAGTGCAGCTTAATGTCCTTATTCAGTAAGGCAAAATCACAGAT 6139
QY 841 ggaagacagattgaaagaacaaattctcatgcatctgtaattgatagtgtactaaag 900
DB 6140 GGAGACAGATTGAAAGGAAGCAATTCATGCAATGTAATGATAGTGTACTATAAAG 6199
QY 901 aggcagctcaaaagacataccagttcaacaaataactatactatgatactgagactct 960
DB 6200 AGGCAGCTCAAAAGACATACAGATTCAACAAATACCTATATGATGATCTGAGGACTCT 6259
QY 961 gaaggaatcttactacgtaggtcaatttgaatagatcaggtcttccaagtatggaaa 1020
DB 6260 GAAGGAATCTTACTACGTAGTAAATTTGTAATAGATCAGGCTTTTCAAGTATGAGAAA 6319

```

```

QY 1021 gtaactaaagtgtcatatcagctcacaaagaacaaatatactcttccacagtggt 1080
DB 6320 GTATCTTAAAGTGCATATCATAGCTCACAGAACCAAGATATTAATCTTTTCCACGTGT 6379
QY 1081 gtagtagctgctgcagatctacacaaactataaaaaacccatagggtctctgtttt 1140
DB 6380 GGAGTAGCTCCCTGCATATCTACCACTATTAATAAACCCATAGGGCTTCTGTGTTTGG 6439
QY 1141 cctcttctcttggaagactggcctgcacattcagtgtaagtggcaacttgacagatcca 1200
DB 6440 CCTCTTTCTTTGGAGACTGGGCTGCTCCATTTCAATGTAATGCGCACTTGACATGATTA 6499
QY 1201 gccagaagaacctgtagcgtgataatagagtgltgltcgaaagtgaactggaataac 1260
DB 6500 GCCAGAAAGAACTGTGGCTGATGATTAATGAGATTGCTGTTGGAAGTGAAGTGAATAC 6559
QY 1261 agtttaatgacagatataatagctcctgcatag-tgaaatgctaatacagttaaaaa 1319
DB 6560 AGTTTAATGACACACTTAATAGCTCCCTGCATATGTTGAATGCTTAATACAGTTAAAAA 6619
QY 1320 cggtaattccctgltcttgatccaacatatacagtggtttacagaacaccctattcagt 1379
DB 6620 CGGTATTTCCCTGGTTGTGATCCACATTTATCAAGTGTACAGAGACCCCTATTATCAT 6679
QY 1380 gtaaaagacacttaagaagat 1401
DB 6680 GTAAGGACACTTTAAAGAAAGT 6701

```

```

RESULT 2
AF193556 12793 bp DNA linear PRI 07-FEB-2000
LOCUS Homo sapiens saccin (SACS) gene, complete cds.
DEFINITION AF193556
ACCESSION AF193556
VERSION AF193556.1 GI:6907041
KEYWORDS human.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 12793)
Engert,J.C., Berube,P., Mercier,J., Dore,C., Lepage,P., Ge.B.,
Bouchard,J.P., Mathieu,J., Melancon,S.B., Schalling,M.,
Lander,E.S., Morgan,K., Hudson,T.J. and Richter,A.
ARSACS, a spastic ataxia common in northeastern Quebec, is caused
by mutations in a new gene encoding an 11.5-kb ORF
Nat. Genet. 24 (2), 120-125 (2000)
JOURNAL 20120709
MEDLINE 2 (bases 1 to 12793)
REFERENCE Engert,J.C., Berube,P., Dore,C., Lepage,P., Ge.B., Hudson,T.J. and
AUTHORS Richter,A.
TITLE Direct Submission
JOURNAL Submitted (08-OCT-1999) Genome Centre, Montreal General Hospital,
1650 Cedar Ave., Montreal, QC H3G 1A4, Canada
FEATURES
source
1..12793
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="13"
/map="between D13S232 and D13S292"
1..12793
/gene="SACS"
/product="sacsln"
1..12793
/gene="SACS"
77..11566
/feature="SACS"
/note="molecular chaperone"
/codon_start=1
/product="sacsln"
/protein_id="AAF31262.1"
/db_xref="GI:6907042"
/translation="MNTFWPGRLLIVQWYPFDENRNPVSWLKMKNLYIHSEDL

```

TLDEMLPIRLPHEGGOCVELILRLRSIVLYLDESENOJDEPLADLYXOKIGGPFVIT
 KKLASIOHPIRRKLEGGONPVSNOIMKMKLOKCONOTSLPPTHKOLAPLRLSTVSD
 TDSSEKEREIIOELAFKRIHNSDDGISYTKLQKGVKHLHAKPADLRLSTVSD
 NPNVLEWTLPLAKFIOLOWTYSCLYKLYEDIEAFLYSHHEVYTOIMLVNLSLNLK
 HSROIGKLNESLSEKEDYVOAKKEIALOQVAGPODQYLKAKTLLVINKHNTLL
 OSSEKMTKLKIMWPAKCEKPPNPNYSGVWMLDCLNCLAPCMQGVGHLIGSSLD
 VHEJHVNLEKRLGIFETRSLASVLYHFWYDWSKSPFDEDDYQOHOILLEYEFL
 MHDHNEEDSRALKPVMYVGTCKCPALQAVYKEIHDLDOPLHNYTKHAKFHO
 LFKVCSIEBELTSHISYQVLYKSDSLDSQESKONHMLNITRWLYSNOIPAS
 PMPVPAHHSKNSKIMKPHHECCQCDKRYVDNLNDESVPIELVHNDIPKATKE
 VLYKVSCLTRILNPNMEGOSGOREPLTVIRKINILEYSDPSEFELVHNDIDANA
 TECSLIDIMRRMADRERLIDGMACHAPMLSPNNSGSDKDFVITRGSILKRGKNG
 EDYKRGKLEGNYSYVHIIDPIINREMEIMDPININHSIKKONGKIMSKK
 OKRLRKPDEKSLPDRNGCOLPLTVNAPYVNGVILRSLFPTQOEAKEYSVSYTY
 NTADYILPDSKCGHRLITQYSKMYKLSKLEBPSKLOLTVYEEFELVHNDIDAN
 ALNTVYLSVLEKAALMKTCSSSNKRLPDEKSCLOLTVYEEFELVHNDIDAN
 LFERGGDAPLEAKSOSKSPBLSOGLQVETCTYKLCMDIGEALFSLSESGS
 RRLGLPGCAVVOLEIDOKMWTVPYHGEVYCYPLRKTGLPVIINCEGATSNK
 KEIMTQDIEGRANTFMHRIYKAYLOVYLSKDLATSGELMLOYAYWPDHFL
 DESVCOGGYEDIEHAGKELKELVDSGTSWMSKVRLEDSYILKRGDGAAPKIF
 LKYTKKTSKCNELSPSVLEKREACKOILLENTPSEKOPSPSVFPNTOIEA
 ELRDPMLFVLEKVEDSGVLRATPCISLSEGLHPLVSLIHEKGRVAKPLID
 GREPPSYQDYNPLILKLYOLGAKMDILOMDMERVSLIHEKGRVAKPLID
 LILSLIDELKLRIDRPAKOPAKYOTRPLPPLKAGESILDMKNGSKRETEMAAD
 LYTLASDOLVILCLOLILNNSHSCGGSVSLAVKELGLKRPVLYVNOLEKAVK
 SVDSITLYOENETNACYKYLEHALONETKSKSIDKLPSEFIVENYVSEKYS
 PHLEAPAYVLOLNKYNKREBELFETVYRGSCVDEVALTESIDORGRKOTHE
 ENFOLCRITISGIMSLPERKOEKREKMYGKILPDTNMLPASKCYNDOPMIV
 KOTKYCHADIPREYVAKVAPKRAKELBERVASHNCPYTLTEREQKRLTISR
 LINAPESEKEMKLELDONDAKAEKICEVDEPRADNVPIDKRAPOGLCAPKYSN
 NOPFEDYVGRIONLQKGEKBNPRTQYQJGDSNYALTDPCPSFGNDINL
 HARVAPGYSISPERGFRDLPDAPROFJEDLYGTERHKLNCYFPRPLNAEPA
 KYSELSVPASOROVONLIDKRSKDAELMEHMKETICEIDKSTGNTLYSVK
 KGTIDGRKROFPAVYSDYSTRKQOLDIPVOYLTMYDESGEANTLWILGNS
 GFSMSKEXKYSIVAKHODITLPPGVAACATHNKYPKRAKPLSLGELPPEH
 VNGHIALDSARNMLRMDGVCYRSOMNNSMIALAPAVELLIOKKRYFOSDPPT
 LSVONDTPLHAYDLKPLSEFPVNRULDOPULCVAKVWCEHDMKRLVYVA
 PNIQSDSHASYITWIIINSTSKNRPFDNLLDDELQHLKANDYNTTKRYVAENVY
 RLKHLLEGNVLVNCDBTANVHCLIDADIPSVYTPADISFLATFESSPTNCHIK
 GFLPRLQOOLNKLHSLKLYDVCYKADKEENITEVEGLDILLTSDLOLFDKRRK
 FLTYHLEIPSKDLPMNLILKYSNILNCKAKVAFIDSSPLDLSVILPREKTK
 CTKKDNKASWESKLNAMHIESVSVQDEBTKRPTDIOYOTLKDMLLPKTPVY
 SANOLVYBEGVLEPLSLMHLIYVPAQSDKVFHAKKACIOALAKLKOSDAP
 LLSCHTANESTSLIKALHAYVOSTPFAEKYLEVDFEALMAYPCNMLHNSOAPD
 KILKSLPYKTSISGRVYSGKTCYVUKRSJPSAEVETKOSSSAFELKILKEL
 YEIVGCVDEDEYVGLKILPLENTSYSLAKLEHLYIKRSLASAELESEKOLPEK
 LESLILIHANRSKLOAKHOFYBTVFVEMFLPDEKTEKILPKHMFHPEH
 MTSWEFLINRGLKLOKSOOLLOKFAKELSVRANTEKMSKELONTYDILHNF
 DDLISGNFLKELSLIPLCRBRAPAEFTIRHPOYOENGTPLIKRGAOYVNRKCOO
 VYLDLISGFLSCPLPEAPLSTRKEDSGSDGPOEOLVYNLNVNLDPLDYKIN
 NINICNTILDEMEVTRAVULRSTIEFSLAEKREFRQOLGAVFVWDMKDLINPC
 VINLNESEDFPYKPLLELGTGFPQHLKILGETDITSRQVENVLSRIFKSESGO
 LDNEMRYKRVYSGFLRSDNSVYRKLDEENVRDATALPODGVYKSLIVPDD
 APHYKSRLOAGNGOVLVOSOCYVYKDEHGFIRKLMPQKRLPRLSTIECKLAR
 EHPKPGACGLSLOGRLOLILSSPOFINGIRIMHNEANDLAPANEKRLICKLAR
 EGKYSCEKLOTTLRYKGFNPRIHRSSTPFAFLKFGNAVILLYIYHDSKIDNPL
 ALMTLKSXTYDNLISOTSLYALMGAGVLYDIAEGDIDGYSVOTYTYVALIY
 PIPATIRYTLMDPMNVFPGVEYGLVYDIAEGDIDGYSVOTYTYVALIYVEB
 NSSFLGKTYQIDIGSEKYSVSLDYKRSREPPSVOSRQASSTPSPPEFLTGR
 SIPLPESEKHSKTSKOSHPKLYKNSLPELILKLEYSVOSQAMKLPESKRIIL
 YLKMDKNPNEANDLANSHEKSEGOOKINREKOPAFONDNRASRPFSSVAGSD
 KYSRPFATYNNMOEANDSHKSEGOOKKECPAGSOTYORFPAVPPFKSVGNFVAR
 RMLROAPNEASAUNDILKANEMWCKYULSKLALIDAVYVRGSDKDYKPTALA
 OKIEKYSOOLBENYHLEAVGVSLKTRIPDILLPFOIDPMDRFTSEVAMRMECT
 ACILIKLEPMOKVY

2256 c 2487 g 3887 t

Query Match	99.28;	Score 1390;	DB 9;	Length 12793;
Best Local Similarity	99.98;	Pred. No. 0;		
Matches 1401; Conservative	0;	Mismatches	0;	Indels 1; Gaps 1;

OY	1	gtgacgtagtaaaactctgaggagcagctcccaaaagcgacacacaaagccttagaagaatactgaccc	60
Db	5300	GTACACAGTAAAACTAAGGAGGACAGTCGCCAAAGCGCACAAAGCCCTTGAAAGATATGCATCC	5359
OY	61	aatctcgttttaacaaactctggcacagaatttggccaaagaagaaatctaccacgcga	120
Db	5360	AATGCTGTGTTTTTACACACCTTGGCACAGATTTTGGCGCAAGAAATTTGACACGCAGA	5419
OY	121	attaagagcactccttaatgcatatccctcttgaaaaggaatctgtgaaagagcttcttcaa	180
Db	5420	ATTAAAGGAGCANTCCTTATGCAATTCCTCTCGAAAGGAATTTGAAAGAGCTTCTTCA	5479
OY	181	aatgcctgaatgacgaagcgacagaaatcgttctgtgttctgacacctagacgacatcca	240
Db	5480	AATGCTGATGATGCAGGAAGCGGACAGAAATCTGTTTGTGTGATCTTAGACACATCCA	5539
OY	241	gttgatgaaatatctgtagtaagctgggcccatttgcgaagcgccgaactctgtgtctac	300
Db	5540	GTTATATGAATATTTGATGATGTAAGTGGGCCCATTTGCGCAAGGGCCGACGACTTGTGTGAC	5599
OY	301	aacaccagcccatcttaacagagaatgactgttagaggaattcagaactctggaaaagcgacg	360
Db	5600	AACACACGACCATTTACAGAGATGATGTTAGAGGAATTCAGAACTTTGGAAAAGCAGC	5659
OY	361	aaaaggagaaatcccttaaaacttggcacgtagtgaatgagatccaattcctgtgtacat	420
Db	5660	AAAGGGGAATCCTTTATTAACCTGCAGATATGGAATAGATTCATATTCGTGATCAT	5719
OY	421	atcacgaacttgcgcattcttlatctcttggcaatgacatccctgttatttttggacctcat	480
Db	5720	ATCACAGATCGCCCATCTTTTATTTCTGGCAATGATCCCTGTAATTTTGTGATCCCAT	5779
OY	481	gcacgaatgacccaaggcgccacatcatctagctcccgagcgactttagagaatttggat	540
Db	5780	GCCAGATATGCACCAGGGGCCACATCATTATGTCGGAGCGATTTTAGAGATTTGGAT	5839
OY	541	gcgaatttaggacacagctctcagaatgtctctgtagcttlatcttggaaaccattttaa	600
Db	5840	GCAGATTTTATAGACACAGTCTCAGATGTTCTGGAATCTTATCTGGAAACCATTTTAA	5899
OY	601	ctggaataatgacacaatgttcagaatttcctctctgttaatgacgaaaatggcaaaagtctcg	660
Db	5900	CTGATATATTCACAAATGTTCAGATTTCTCTTCGTAATGACGAATGGCAAAAGTTTTCG	5959
OY	661	gaattctcgtctgttccagaatcagaacgaagaatgtgcgaatcctttggcaaaaactcgcg	720
Db	5960	GAAATTTCTGTCGTTCCAGCATATGACAGAAATGTCACAAATCTTTGACGCAAACTGGC	6019
OY	721	tcaagatgggaggaacttccaacttcttcttaacatgacatggaaaataattctcatctgtgaa	780
Db	6020	TCAGATGGGAGAGACTTCTTAATGTTTCTTAATCANTGAAAATTTCTATTTTGTGAA	6079
OY	781	atagataagagtagcttggagctcctaaatgtgtctgtatctcagtaaaaggcnaaaatcaagat	840
Db	6080	ATAATATTAAGAGTACTCGAGCTCTTAATGTGCTGATTTCTAGTAAAGCGCAAAATCACAGAT	6139
OY	841	ggagaacgatttgaagaagaaacatctcatgcatctgtaattgataagtgttacctaaagaag	900
Db	6140	GGAGACAGATTGAAAGGAAACAAATTTTCATATGCACTCTGTAATTTGATATGTTACTTAAAG	6199
OY	901	aggcagactccaagaacatcacagcttcaacaaataacctatactatgtagtactggagactct	960
Db	6200	AGGCAGCTCAAAGCANTACCAAGTTTCAACAAATTAACCTAATGATGATACGTAGAGAGCTCT	6259
OY	961	gaaggaaactcttactacgttggctcaatttctgtaataagatcaggcttctcaagtatggagaa	1020
Db	6260	GAAAGAAATCTTACTACGTGCTGAATTTGTAATATGATCAGCTTTTCAAGATATGAGAAA	6319
OY	1021	gtactctaaagtgtcatatcatgctcaacgaagaacaaatatattactcttttcccacgtgt	1080
Db	6320	GTATCTTAAAGTGCATATATAGCTCCAAAGAACCAAGATATTTACCTTTTCCACAGTGCT	6379

```

QY 1081 ggaagtagctgcctcattactcaacaatacaaaacccatagggccttcgtttttg 1140
Db 6380 GGAGTAGCTCCCTGCATTAATCACTATATAAAACCCATAGGCGCTTCTGTTTTTG 6439
QY 1141 cctcttcttcttgagagctggctgccttcaatgtaagtgccacttgcactgattca 1200
Db 6440 CCTCTTCTTCTTGAGAGCTGGCGCTCCATTCATGTAATGAGCCACTTGCACCTGATTC 6499
QY 1201 gccagaagagacactggtggctgtagtaataatgagtggtgttctgaaagtacgaaataac 1260
Db 6500 GCCAGAGAGAACCTGTGGCTGATGATATGAGAGTGGTGTGCAAGTGAAGTGAATAC 6559
QY 1261 agttaatgacagcatatagctcctgcatacg-tgaatgtctaatacagttcaaaaaa 1319
Db 6560 AGTTATATGACACATTAATATAGCTCCTGCATATGTTGAATTGTAATATCAAGTTAAAAA 6619
QY 1320 cggatattccctcgttctgatacaacattcagtggttaacagaacacctatcatgtt 1379
Db 6620 CGGATATTCCTCTGTTCTGATCCACATATACAGTGTACAGAACACCCCTATTCATGTT 6679
QY 1380 gtaaaagacacttaagaagt 1401
Db 6680 GTAAAGCACACTTTAAGAAAGT 6701

RESULT 3
AL157766/c 92693 bp DNA linear PRI 11-APR-2001
LOCUS Human DNA sequence from clone RP11-40020 on chromosome
DEFINITION 13q12.11-12.2, complete sequence.
ACCESSION AL157766
VERSION AL157766.9 GI:13620292
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 92693)
REFERENCE Submitted (11-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
AUTHORS CB10 15A, UK. E-mail enquiries: humquery@sanger.ac.uk
TITLE requests: clonerequests@sanger.ac.uk
JOURNAL On Apr 12, 2001 this sequence version replaced gi:12709868.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submissions
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em, EMBL; Sw,
SWISSPROT; Tr, TrEMBL; Wp, WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormep This sequence
was generated from part of bacterial clone contigs of human
chromosome 13, constructed by the Sanger Centre Chromosome 13
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr13
RP11-40020 is from the library RPCI-11.1 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACE3.6
IMPORTANT: This sequence is not the entire insert of clone
RP11-40020. It may be shorter because we sequence overlapping
sections only once, except for a 100 base overlap.

```

```

FEATURES
Source
The true left end of clone RP11-760M1 is at 92594 in this sequence.
The true right end of clone RP11-72P19 is at 100 in this sequence.
location/Qualifiers
1..92693
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="13"
/map="q12.11-12.2"
/clone="RP11-40020"
/cclone_11b="RPCI-11.1"
2390..2485
/note="MIR repeat: matches 81..192 of consensus"
2562..2673
/note="MIR repeat: matches 68..183 of consensus"
3896..4201
/note="AluY repeat: matches 3..308 of consensus"
5122..5397
/note="MER46C repeat: matches 1..286 of consensus"
18986..19294
/note="AluY repeat: matches 1..310 of consensus"
19644..19873
/note="MER46A repeat: matches 1..235 of consensus"
20613..20912
/note="AluSg1 repeat: matches 1..300 of consensus"
23342..23651
/note="AluSg1 repeat: matches 1..309 of consensus"
24769..24891
/note="L2 repeat: matches 2554..2662 of consensus"
23871..26011
/note="L2 repeat: matches 2356..2495 of consensus"
26033..26109
/note="L2 repeat: matches 2601..2688 of consensus"
26245..26344
/note="L2 repeat: matches 2154..2255 of consensus"
26938..27096
/note="MIR repeat: matches 3..175 of consensus"
27150..27653
/note="L2 repeat: matches 1063..1644 of consensus"
28522..28891
/note="MER1B repeat: matches 1..364 of consensus"
29447..29834
/note="L1ME3A repeat: matches 5787..6164 of consensus"
36098..36415
/note="AluSx repeat: matches 1..308 of consensus"
37202..37414
/note="MIR repeat: matches 22..262 of consensus"
37963..38254
/note="AluSg repeat: matches 9..301 of consensus"
38703..39008
/note="AluSg repeat: matches 1..306 of consensus"
39790..40093
/note="AluSx repeat: matches 1..304 of consensus"
40126..40416
/note="AluSg repeat: matches 1..292 of consensus"
40444..40733
/note="AluSg repeat: matches 1..292 of consensus"
41322..41405
/note="Single clone region. Assembly confirmed by
restriction digest data."
41541..41788
/note="AluSg repeat: matches 1..248 of consensus"
44790..45101
/note="AluSg repeat: matches 1..313 of consensus"
45261..45312
/note="13 copies 4 mer tggc 888 conserved"
45899..46206
/note="AluY repeat: matches 1..307 of consensus"
46754..47052
/note="AluY repeat: matches 1..298 of consensus"
47067..47365
/note="AluY repeat: matches 1..299 of consensus"
47477..47873
/note="L1MA10 repeat: matches 5950..6322 of consensus"

```



```
repeat_region 47889..48229
                /note="AluXs repeat: matches 1..312 of consensus"
repeat_region 49168..49212 repeat: matches 85..126 of consensus"
                /note="Alu repeat: matches 85..126 of consensus"
repeat_region 49620..49693
                /note="L2 repeat: matches 1685..1757 of consensus"
repeat_region 50704..51032
                /note="AluSc repeat: matches 1..308 of consensus"
misc_feature 52204..53009
                /note="CpG Island"
                /evidence="not_experimental"
repeat_region 53978..54137
                /note="AluXs repeat: matches 1..160 of consensus"
repeat_region 54179..54511
                /note="LMB6 repeat: matches 5822..6172 of consensus"
repeat_region 53685..55949
                /note="AluXs repeat: matches 1..305 of consensus"
repeat_region 57331..57390
                /note="AluXs repeat: matches 1..305 of consensus"
repeat_region 57357..57392
                /note="30 copies 2 mer ga 75% conserved"
repeat_region 57753..57930
                /note="9 copies 4 mer gaga 91% conserved"
repeat_region 58260..58389
                /note="MIR repeat: matches 82..262 of consensus"
repeat_region 58564..58611
                /note="MIR repeat: matches 2..153 of consensus"
repeat_region 59350..59533
                /note="24 copies 2 mer ca 93% conserved"
repeat_region 59992..60223
                /note="AluXq repeat: matches 129..313 of consensus"
repeat_region 61036..61144
                /note="AluJo repeat: matches 85..299 of consensus"
repeat_region 62008..62187
                /note="L2 repeat: matches 2581..2656 of consensus"
repeat_region 62188..62316
                /note="TIGER1 repeat: matches 2238..2418 of consensus"
repeat_region 62330..62363
                /note="AluY repeat: matches 1..129 of consensus"
repeat_region 62362..62565
                /note="Alu repeat: matches 261..294 of consensus"
repeat_region 62566..62865
                /note="TIGER1 repeat: matches 1586..1787 of consensus"
repeat_region 62866..64385
                /note="AluXq repeat: matches 1..302 of consensus"
repeat_region 64386..64694
                /note="TIGER1 repeat: matches 46..1586 of consensus"
repeat_region 64695..64713
                /note="AluY repeat: matches 1..306 of consensus"
repeat_region 65068..65395
                /note="TIGER1 repeat: matches 29..46 of consensus"
repeat_region 65396..65569
                /note="LIPB2 repeat: matches 5405..5733 of consensus"
repeat_region 65571..65640
                /note="AluY repeat: matches 136..309 of consensus"
repeat_region 65696..65717
                /note="LIPB2 repeat: matches 5728..5791 of consensus"
repeat_region 65725..66096
                /note="11 copies 2 mer ta 100% conserved"
repeat_region 66371..66410
                /note="LIPB2 repeat: matches 5789..6155 of consensus"
repeat_region 67586..67886
                /note="10 copies 4 mer tgtg 82% conserved"
repeat_region 69748..69930
                /note="Aluub repeat: matches 1..299 of consensus"
repeat_region 70957..71267
                /note="MIR repeat: matches 6..248 of consensus"
repeat_region 71279..71413
                /note="AluY repeat: matches 1..311 of consensus"
repeat_region 71411..71737
                /note="MER21B repeat: matches 548..680 of consensus"
repeat_region 71780..72075
                /note="MER31A repeat: matches 47..485 of consensus"
repeat_region 71780..72075
                /note="AluXs repeat: matches 1..295 of consensus"
```

```
repeat_region 72145..72256
                /note="MER31-internal repeat: matches 42..175 of
                consensus"
repeat_region 72454..72865
                /note="MER31-internal repeat: matches 332..739 of
                consensus"
repeat_region 72873..73249
                /note="MER31-internal repeat: matches 883..1261 of
                consensus"

Query Match 99.2%: Score 1390; DB 9; Length 92693;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1401; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 gtagcaataaactcaggagcagtcctccaaagcgacacaaagccttagaagaatcgtcc 60
DB 13610 GTAGCACTAAACCTAGAGAGCGTCCCAAGGACACAAAGCCTTAGAAGATGATGCATCC 13551

QY 61 aatgctgctttacacacctgycacagaatttggcagaagaanaattgaccagaga 120
DB 13550 AATGCTGCTTTACACACACTTGGCAGCAAAATTTGGCAGAAAGAAAATTGACCAGCAGA 13491

QY 121 attaagacatcccttaatgcatatccctctgaaagaanaatgtaagaagcttctcoa 180
DB 13490 ATTAAGACATCCTTAATGCATATCTTGAAAGAAAAGTAAGTAAGACCTTCTTCAA 13431

QY 181 aatgctgatgacgaagcgacagaagaatcgtttgtgttgatccttagacagatcca 240
DB 13430 AATGCTGATGACGAAGCGACAGAGAATCGTTTGTGTGATCCTTAGACACATCCA 13371

QY 241 gttgataaatattgtagtaagtggcccatctgcaagggccagcactgtgtgtac 300
DB 13370 GTTGATAGAAATATTGATGATGAATGGGCCCATTTGCAAGGGCCAGCATTGTGTGTAC 13311

QY 301 aacacaccagccattacagaagaatgatalgtagaagaaattcagaattctggaanaagcag 360
DB 13310 AACACACAGCCATTACAGAAAGATGATGTAGAGAAATTCAGAAATTCGGAAGAGCAGC 13251

QY 361 aaagaaggaaatccttataaacctgacagatgataagataatcctcgtgtatcat 420
DB 13250 AAAGAGGAAATCCTTATTAACAGGACAGATGGAATAGCATTCATTCGTGTATCAT 13191

QY 421 atcacagactggccatcttattccttcgcaagatgacatccgtgtatttggatcccat 480
DB 13190 ATCACAGACTGCCATCTTTATTTCGTGCAATACATCCTGTATTTTGTGATCCTCAT 13131

QY 481 gccagatcgcaccaggggcccacatccatagtcgccgagcagatgtaagaattgtag 540
DB 13130 GCCAGATATGCACAGGGGCCACATCCATTATGTCCGAGCCAGCTTTAGAGATTTGGAT 13071

QY 541 gcagattttagcacagacttccagaatgcttgatcttatactcgggaaccatttaaa 600
DB 13070 GCAGATTTTAGCACAGACTTCTCAGATGCTTGATCTTTATTCGGGAACCATTTTAAA 13011

QY 601 ctgataaattgcacaatgcttcaagattcccttccttgtaatgcaagaatggcaaaagtttcg 660
DB 13010 CTGATAAATTGCACAATGTTTCAAGATTTCTTCTTGTAATGCAGAAATGGCAAAAGTTTCG 12951

QY 661 gaaattcgtctgtccagcatcgcagacaagaatgctcagaatctttggaacaaatcgcgc 720
DB 12950 GAAATTCGCTGTCCAGCATCAGACAGAAATGCTTCAGAAATCTTTGGCAAAATCGCGC 12891

QY 721 tcagatgggacagaacttcaatgcttcttaataatcaatgtaaaatcttcaatttgtaa 780
DB 12890 TCAGATGGGACAGAACTTCAATGTTTCTTAATACATGGAATAAATTTCAATTTGTGAA 12831

QY 781 atagataagatgactggagctctaaatgctgctgattacagtaaaaggcaaaatcaagat 840
DB 12830 ATAGATAGAGTACTGGAAGCTTAAATGTGCTGATTTCAGTAAAGGCAAAATACAGAT 12771

QY 841 ggaagacagatgtaaaagaaacaaattcatgcatcgttaattgataatgtaattgtaactaaag 900
DB 12770 GGACACAGATTGAAAGAAAGAAACAAATTCATGCAATCTGTAATGTGATGATTAGTAAAG 12711
```

OY	901	aaggagcctcaagaacatccagttcaacaataaaccttactgatatgaagacacct	960
Db	12710	AAGGAGCTCAAAAGACATACCGATTACAATAAACCTTACTATGTGAATCAGAGACCTT	12651
OY	961	gaaggaatcttactcacgtgcctaatttgtaaatagatcagcctttccaagttagaaa	1020
Db	12650	GAAAGAATCTTTACTACGTGGCTAATTTTGTAATAGATCAGCCTTTTCAAAGTATGGAGAAA	12591
OY	1021	gtatctaaagtgtcatatcagctcacagaaccaagatatctcttttcccacytgt	1080
Db	12590	GTATCTAAAGTGTCATATCATAGCTCACAAAGAACCAAGATATTCATCTTTCCACAGTGGT	12531
OY	1081	ggagagtcgtcgctgatctaccacaacacataaaaaaaccocataaggccttcglttttg	1140
Db	12530	GGAGTAGCTGCCTGGATTACTCATACACATATAAAAAACCACATAGGCGCTTCGTGTTTTTGG	12471
OY	1141	cctcttctcttggagacgcggcctgcacattcaatgtaatgagcacttggacatgattca	1200
Db	12470	CCTCTTCTTTTGGAGACCTGGCGCTGCCATTTCATGTGAATGGCCACTTTGCACTGCATTTCA	12411
OY	1201	gccagaagaacccgtgtagatgatalaataatggagttggtgtcgaagtgactggaataac	1260
Db	12410	GCCAGAAGAAGACCTGTGGCGATGATATATGAGATGTGGTGTTGGAAGTACGATGGAATTAAC	12351
OY	1261	agtttaatgaacagcatataatgctctctycatatg-tgaattgctaatcagittaaaaaaa	1319
Db	12350	AGTTTAAAGACACAGCTTATATAGCTCTCCATATGTGTGAATTCCTATATACAGTATAAAAAA	12291
OY	1320	cggatctccctcgtgctcgtatccacaacatataagtggttacagaagaacccttatggtt	1379
Db	12290	CGGATTTTCCTCGTTCGTTCGATTCACCAATTTATTCAGTGTTCACAGAACCCCTTATTCATGTT	12231
OY	1380	gtaaaagacactttaagaagt 1401	
Db	12230	GTAAGGACACTTTTAAAGAGT 12209	
RESULT 4			
LOCUS	AC079761/c	99819 bp DNA linear HTG 10-SEP-2000	
DEFINITION	Homo sapiens chromosome UNK clone RP11-143617, *** SEQUENCING IN		
PROGRESS	***, 44 unordered pieces.		
ACCESSION	AC079761		
VERSION	AC079761.1 GI:10047966		
KEYWORDS	HTG; HTGS_PHASE1.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
TITLE	1 (bases 1 to 99819)		
JOURNAL	Waterston,R.H.		
REFERENCE	The sequence of Homo sapiens clone		
AUTHORS	Unpublished		
TITLE	2 (bases 1 to 99819)		
JOURNAL	Waterston,R.H.		
COMMENT	Direct Submission		
	Submitted (10-SEP-2000) Genome Sequencing Center, Washington		
	University School of Medicine, 4444 Forest Park Parkway, St. Louis,		
	MO 63108, USA		
	----- Genome Center -----		
	Center: Washington University Genome Sequencing Center		
	Web site:http://genome.wustl.edu/gsc/index.shtml		
	Project Information -----		
	* NOTE: This is a 'working draft' sequence. It currently		
	* consists of 44 contigs. The true order of the pieces		
	* is not known and their order in this sequence record is		
	* arbitrary. Gaps between the contigs are represented as		
	* runs of N, but the exact sizes of the gaps are unknown.		
	* This record will be updated with the finished sequence		
	* as soon as it is available and the accession number will		
	* be preserved.		

[illegible]

QY 721 tcagatgagggcagaactctcaatgtcttcaatcacaatggaataatcttctattgtgaa 780
|||||
Db 87476 TCAGATGAGGGCAGACACTTCTTAATGTTTCTTAATCACAATGGAATAAATTTCTATTGTGAA 87417
QY 781 atagataagagtaactgagctcctaataatgctglatcagtaaaaggcacaatcacagat 840
|||||
Db 87416 ATGATGATAGAGTACTGAGAGCTTAATGTCGTATTCATTAAGGCAAAATCACAGAT 87357
QY 841 ggaagacagattggaaggaaggaacaatttcagtcactgttaattatagtgcttactaaag 900
|||||
Db 87356 GGAAGACGATTGAAAAGGAAGCAATTTCAATGCTGTAATGTAATGTTGTTACTTAAAG 87297
QY 901 aggcagctcacaagacataccagcttcaacaataccatactatggaactc 960
|||||
Db 87296 AGGCAGCTCAAAACATATACAGATATCAAAATGSCATATCTATGATATGTAAGACTCT 87237
QY 961 gaaaggaatc-ttaactagctgcttaattgttaataagtcagagcttttcaagatgagaa 1019
|||||
Db 87236 GAAAGAAATGTAAGTACAGCTGAGCTAATTTGTAATGATCAGGCTTCAAGATGAGAA 87177
QY 1020 agtatctaaagtgctatcatcagctcacaagaaccaaagatactcttcccaagctg 1079
|||||
Db 87176 AGTATCTAAAAGTGTATATACAGCTCACAAGAACCAAGATATTACTCTTTTCACAGCTGG 87117
QY 1080 tggagtagctgctgcatcactcacaactataaaacccatagggccttctgtttt 1139
|||||
Db 87116 TGGAGTAGCTGCTGCTGCTATCTCACAACATATAAAAAACCCATAGGCTTCTGTTTTT 87057
QY 1140 gctcttctcttgagagctgagctgctcattcagtgaatgagcaacttgcactgagttc 1199
|||||
Db 87056 GCTCTCTTCTTTGAGAGCTGAGCTGAGCTGCTATTCATGTAAGGCTTCTGCTGATTC 86997
QY 1200 agcagaagaaggaactctgagctgagatgaatgagtggttcgaagtagcagagataa 1259
|||||
Db 86996 AGCCAGAAAGAACCTGTGAGCTGATGTAATGAGAGTGGTGGTGAAGTACAGTGAATAA 86937
QY 1260 cagtttaatgacagcattaaatagctcctgcatatg-tgaattgctaatacagttaa 1318
|||||
Db 86936 CAGTTTAATGACAGCAATTATAGCTCTGATATGTTGATTTGATTTGATTAAGTTAA 86877
QY 1319 acggtatctccctgctgctgtaaccacaatatacagtgtaacagaacacccattcatgt 1378
|||||
Db 86876 ACGGTATTTTCCCTGCTGCTGATCCACATATATGATGTTACAGAACACCCCTATTTCATGT 86817
QY 1379 tgfataaggaactttaaagaagt 1401
|||||
Db 86816 TGTAAAGACACTTTAAAGAGT 86794

RESULT 5
AX119933 11492 bp DNA linear PAT 11-MAY-2001
LOCUS AX119933
DEFINITION Sequence 3 from Patent WO0129266.
ACCESSION AX119933
VERSION AX119933.1 GI:14036679
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 11492)
AUTHORS Hudson,T.J., Engert,J. and Richter,A.
TITLE Identification of aarsac mutations and methods of use therefor
JOURNAL Patent: WO 0129266-A 3 26-APR-2001;
MCGILL UNIVERSITY (CA); Hopital Sainte-Justine (CA)
FEATURES
source 1..11492
/organism="Mus musculus"
/db_xref="taxon:10090"

BASE COUNT 3599 a 2280 c 2387 g 3226 t
ORIGIN

Query Match 83.8%; Score 1174; DB 6; Length 11492;
Best Local Similarity 90.3%; Pred. No. 8, 8e-265;
Matches 1266; Conservative 0; Mismatches 135; Indels 1; Gaps 1;
QY 1 gtagcagtaaaactaggaagcagtcaccaagcgacacaagccttagaagaatgcatcc 60
|||||
Db 5223 GTAGCTGTAAAACTTGCTGCAATACCAAGAGCAATTAAGCATTAAGATATGATCATCC 5282
QY 61 aatgctgttttcaaacacttgcagacgaattgggcagaataaaatggaccacaga 120
|||||
Db 5283 AACATCTGTTTCAAGCTCTAGGTACAGAAATTTGGCGAGAAAGAAAGTACACAGCA 5342
QY 121 atcaagagcaltccatlaatgcatatccctctgtaaaagaaatgltgaaagagcttctca 180
|||||
Db 5343 ATTAAGAGCATTTCTCATATGCTTATCCTTCGAAAAGAAATGCTGAAGAGCTTCTTCAA 5402
QY 181 aatgctgataagcaagcgacagaatcgtttgtgttgccttagacagcatcca 240
|||||
Db 5403 AATGCTGATGATGCAAAAGCCACAGAGATCTGCTTGTGTGATCTAGACAGCATCTCT 5462
QY 241 gttgataagaatatttgataagtggtggcccatctgcaagggccagacttggctgac 300
|||||
Db 5463 GTTGACCGAATATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5522
QY 301 aacacacagcatttcaagagataatgataaggaatcagaatccttgaaagagcag 360
|||||
Db 5523 AACACACAGCATTTTACAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5582
QY 361 aaagaggaatccctataaaactgaaactgataagaaatgaaatcgttgcattcat 420
|||||
Db 5583 AAAGAAAGGAATCTTGCAAAACAGACATTAATGATGATGATGATGATGATGATGATGAT 5642
QY 421 atacagagcagccatcttattcttgcaagaatccctgtaatttttggctccat 480
|||||
Db 5643 ATTACAGACTGCCCTTCTTTATTTCTGCAATGATGATGATGATGATGATGATGATGATGAT 5702
QY 481 gccagatataccacagagggccacatccatagtcaccgagcagatglttaagaattgag 540
|||||
Db 5703 GCCAGATATGACACAGAGGACCATCATGATGATGATGATGATGATGATGATGATGATGAT 5762
QY 541 gccagatttaggaacacagttcagatgttcagatcccttattctgggaacacatttaa 600
|||||
Db 5763 GCAGACTTTAGAACCCAGCTCTCAGATGTTCTGATCTGATCTGATCTGATCTGATCTGAT 5822
QY 601 ctgataatagtcacaatgttcacatttccctctcgtgaatgcagaatgtggaagagttt 660
|||||
Db 5823 CTGACATATGTAACATGTTTATGATTTCTCTGTAATGCAAGATGACAAAGTTTCA 5882
QY 661 gaatttgcctgctcagcatcagacagaatggtccagatccttctggacaactgcgc 720
|||||
Db 5883 GAAATTTCTTCGTTCCATCATATCAGACAGATGCTCCAGAAATCTTTGGCAAGTTACGG 5942
QY 721 tcagatgagggcagaactcttaattgttcttaacatcagtaagaaatcttctattgtgaa 780
|||||
Db 5943 TCTGATGAGGGCAGAACCTTAAATGTTTCTCAACACATGAGAAATATCTATTGTGAA 6002
QY 781 atagataagagtaactgagctcctaataatgctglatcagtaaaaggcacaatcacagat 840
|||||
Db 6003 ATGATGATAGAGCTCAGAGGCTGGAATGCTCTTATTCATTAAGCAAGACTGACTGT 6062
QY 841 ggaagacagattggaaggaaggaacaatttcagtcactgttaattatagtgcttactaaag 900
|||||
Db 6063 GGAAGACGATTGAAAAGGAAGCAATTTCCACGCTCTGTAATTAACAGTGTACTTAAAG 6122
QY 901 aggcagctcacaagacataccagcttcaacaataccatactatggaactc 960
|||||
Db 6123 AGACAGCTCAAGACATATACAGATTAACAAATTAACCTACATATGATGATGATGATGATGAT 6182
QY 961 gaaaggaatcttactagctggtcctaattgttaataagtcaggtcttcaagatagggaa 1020
|||||
Db 6183 GAAAGAAATCTGAACATGCTATCTCTGTAATGATGATGATGATGATGATGATGATGATGAT 6242
QY 1021 gtatctaaagtgctatcagctcacaagaacaaagatatactcttctccagctggt 1080

Db 5284 AACATCTGTTTACAGCTCTAGTACAGAAATTTGGCCAGAAAAGAACTGACACAGACA 5343
 Qy 121 atlaagaagacatccatgaatcctcctcgaagaagaatgttgaagaagctcttca 180
 Db 5344 ATTAAGAGCATTTCTCATATGCTCTCTCGAAGAAAGAAATGCTGAAGAGCTTTTCA 5403
 Qy 181 aatgctgatgatgcaaaagcagcagaatcgtttgttgccttagacacacatca 240
 Db 5404 AATGCTGATGATGCAAAAGCCAGAGATCTGCTTGTGTTGATCTACAGCATCTCT 5463
 Qy 241 gttgatagaatatttgatagaatgtggcccatctgcaaggccagcacttctgtac 300
 Db 5464 GTTGACCGAATATTGATGATGATGATGATGATGATGATGATGATGATGATGATG 5523
 Qy 301 aacaacacacacatcttcaagaagaatgttgaagaatcagaatccttgaagaagcag 360
 Db 5524 AACACACAGCATTTTACAGAGATGATGATGATGATGATGATGATGATGATGATGAT 5583
 Qy 361 aaagaaggaatccctataaactgacagatagatagatcaatctgtatcat 420
 Db 5584 AAGAGAGGAAATCTCTGCAAAAACAGACATTTGATGATGATGATGATGATGATGAT 5643
 Qy 421 atcacagctgcccacatcttatttctgcaatgacatcctgtatcttctcat 480
 Db 5644 ATTACAGACTGCTCTCTTATTTCTGCAATGACATCTGCTGCTGCTGCTGCTGCTG 5703
 Qy 481 gccagatagacacagggccacatcatcattgctccggaagcagatgtttagagattg 540
 Db 5704 GCCAGATATGACACAGAGACCATGATGATGATGATGATGATGATGATGATGATGAT 5763
 Qy 541 gcaagattttagagacagatctcagatgtctcagatcttcttcttcttcttcttct 600
 Db 5764 GCAGACTTTAGAACCCAGTCTCAGATGTTCTGATCTGATCTGATCTGATCTGAT 5823
 Qy 601 ctggatgaattgcccattgttcgattctctcctcctcgaatgcaagaatgcaagaattg 660
 Db 5824 CTGCAAAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5883
 Qy 661 gaattctgtctcctcagacatcagacagatgtctcagacatcttctgcaagaagcgc 720
 Db 5884 GAAATTTCTCTGCTCTCATCATCATGACAGATGCTCCAGATCTTTTGACAAAGTTAC 5943
 Qy 721 tcaagatgggagacacttcttcttcttcttcttcttcttcttcttcttcttcttct 780
 Db 5944 TCTGATGGGGGAGAACTTCTTAATGTTCTCAACCATGAGAAATATCTATTGTTGA 6003
 Qy 781 atagataagagacttgagctcttcttcttcttcttcttcttcttcttcttcttctt 840
 Db 6004 ATGATTAAGGCCACAGAGAGTGTGAATGTCTTATTTCAATTAAGGCAAGATCACTG 6063
 Qy 841 ggaagacagatggaagaagaacacatcttcttcttcttcttcttcttcttcttcttct 900
 Db 6064 GGAGACGATTTGAAGGAAGCAATTTCCACGCTCTGTAAATTGACAGTGTACTTAAAG 6123
 Qy 901 aggcagctcaagaac 960
 Db 6124 AGACACTCTAAGGACATACACAGTTCAACAAATTAACCTACATGATGATGATGATGAT 6183
 Qy 961 gaaggaatcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 1020
 Db 6184 GAAGGAATCTTGACACATGCTCATCTGTAAATGATGATGATGATGATGATGATGAT 6243
 Qy 1021 gtaataaagatgcatcagctcagaagaacacacacacacacacacacacacacacac 1080
 Db 6244 GATTCAGAGATGATATATCAGCTCAGACAGAACCAAGATATACCTTTTCCACAGTGT 6303
 Qy 1081 ggaagatgctgctcattctcacaacataaanaacacacacacacacacacacacac 1140
 Db 6304 GGAGTGGACGCTGATTAATCAATTAATTAATTAATTAATTAATTAATTAATTAAT 6363
 Qy 1141 cctctcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 1200
 Db 6364 CCTCTCTCTTTTGGAGACAGGGCTGCAATTCATGTAATGGCCACCTTGTCTAGATTC 6423

Qy 1201 gccagaagacacctgtgagcgtgatgataatgagatgtgtgtctgaaagtgaatgaac 1260
 Db 6424 GCCAGAAGAAATCTTGAGGCTGATGATGATGATGATGATGATGATGATGATGATGAT 6483
 Qy 1261 agttaatgacacatataatgactcctgcatatg-tgaattgctaataacagttaaaaa 1319
 Db 6484 AGTTAATGACACATTAATGACACCTGATGATGATGATGATGATGATGATGATGAT 6543
 Qy 1320 cggatttccctgtgtctgataacatataatgatttgaagaagacccctattcatt 1379
 Db 6544 CGGATTTTCCCTGCTGTCACCAATTAATGATGATGATGATGATGATGATGATGAT 6603
 Qy 1380 gtaagagacacatataaagaat 1401
 Db 6604 GTTAAGGACACATTAAGAAGT 6625

RESULT 7
 AB006708
 LOCUS 78844 bp DNA linear PLN 27-DEC-2000
 DEPOSITION Arabidopsis thaliana genomic DNA, chromosome 5, P1 clone: MYJ24.
 ACCESSION AB006708
 VERSION AB006708.1 GI:2351073
 KEYWORDS
 SOURCE Arabidopsis thaliana (strain: Columbia) DNA, clone: lib: Mitsu P1 clone: MYJ24.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE
 AUTHORS Kotani, H., Nakamura, Y., Sato, S., Kaneko, T., Asamizu, E., Miyajima, N., and Tabata, S.
 TITLE Structural analysis of Arabidopsis thaliana chromosome 5. II. Sequence features of the regions of 1,044,062 bp covered by thirteen physically assigned P1 clones
 JOURNAL DNA Res. 4 (4), 291-300 (1997)
 MEDLINE 98069011
 REFERENCE 2 (bases 1 to 78844)
 AUTHORS Nakamura, Y.
 TITLE Direct Submission
 JOURNAL Submitted (22-AUG-1997) Yasukazu Nakamura, Kazusa DNA Research Institute, Department of Plant Gene Research, 1532-3, Yana, Kisarazu, Chiba 292-0812, Japan (E-mail: ynakam@kazusa.or.jp, Tel: 81-438-52-3935, Fax: 81-438-52-3934)

COMMENT

Address for correspondence: kaos@kazusa.or.jp
 For the latest information on annotation of this clone, please see <http://www.kazusa.or.jp/kaos/cgi-bin/sgd-graph.cgi?c=MYJ24>
 Genes with similarity to proteins in the databases are described in 'product' or 'note' qualifiers. Genes that have no significant protein similarity are described as 'unknown protein'.
 The software programs used to predict genes include: Grail (Informatics Group, Oak Ridge National Laboratory, <http://compbio.ornl.gov/Grail-1.3/>), GENSCAN (Chris Burge, MIT, <http://CCR-081.mit.edu/GENSCAN.html>), NetGene (S.M. Hebsgaard, et al., CBS, Technical University of Denmark, <http://www.cbs.dtu.dk/services/Netgene2/>) and SplicePredictor (Volker Brendel, Stanford University, <http://genome1.zoool.lasstate.edu/cgi-bin/sp.cgi>).
 Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, Washington University School of Medicine, St. Louis, <http://genome.wustl.edu/eddy/tRNAscan-SE/>).
 This sequence may not be the entire insert of this clone. It may be shorter because we remove overlaps between neighboring submissions. The 5' clone is T2007 and the 3' clone is MKD15.

FEATURES

source
 1..78844
 /organism="Arabidopsis thaliana"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /chromosome="5"
 /clone="MYJ24"

```
/clone_lib="Mitsui P1"
join(8176..8619,8919..9178,9576..9789,9900..9980,
10476..10568,10676..10762,10855..10955,11061..11094,
11261..11335,11485..11607)
/note="gene_id:MYJ24.1"
/codon_start=1
/evidence=not_experimental
/product="2-Isopropylmalate synthase-like protein"
/protein_id="BAB09819.1"
/db_xref="GI:9759360"
/translation="MASLLLTSSSMITTSQCRSWLNSGLPIGSGFSLNLRPYDKAT
LTPVSCSASAKVATSDTLKPIMERREPYIPKPLPHKNVRLDTLTDGSESPAA
LPPOKLEIARQLAKLAVDIMEYGFVSSEEEEAIKITAKITGVNDEDETCVPIC
GIARCKKRIEATWEALAKYAKPRVWLFTSTSEIHMKYIKKTEVIMAVYKVA
KSLGFPDIQCGEDGGRTEKDFCTKIGSRIGATNGFADYVIGIMQOBEELAY
VIENTPGADDYFAIHCHNDLVATANTISGICAGAROVEITTINGESGNAPLEAY
VWALCKRGSLEMDGVYKIDSQIMATSKMVOEHTMYVQPHKPIVGDGCFVIESGIH
OVALKRNSTYEILSPEDYGVKSSENSIVLGLSGRHAVKDLKELGEISDEKEND
IFSRYRELTCKKRITPDADLKALVNGAEISSEKLNKSGINDLMSSPQISAVV"
join(16792..17283,17380..17682)
/note="gene_id:MYJ24.2"
/codon_start=1
/evidence=not_experimental
/product="senescence-associated protein 5-like protein"
/protein_id="BAB09820.1"
/db_xref="GI:9759361"
/translation="MLRLSNAVITTNALLALIGLALSESVYVYQGSQCORPVON
PIVITALLFTSSIGLIALVGSHTITLYLEFLSLILLVLSVFFLVNIPAG
KALSGRIGNVATQYQNMIGNHFLGRKNWEGITKIDSRSVRCRGPRIIDFDSKHL
SNVOFGCCRPPECGFESKNAMWVATPATTAIIGDKAMSTOROLCYACBSCKIG
VLKIGIKRWRITIVNLLILVFLYSCGCCYKKNRPKRRRF"
complement(join(19438..19588,19681..19757,19838..20220,
20533..20633,20839..20903))
/note="emb|CAB62636.1"
/gene_id="MYJ24.3"
similar to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB09821.1"
/db_xref="GI:9759362"
/translation="MSSSLTSGSTVSSSFIAFSKPSLYRNSKTSILDFRNVSRFK
TYKCIYDSSYGVGNVPTFPRTRVMDPYKRLGVSPYASEEETWASRNFLQOYAGHSE
ESLEGAFELKIMSSFIKRRKTKINLSKLIKVEESPMLKALDVEEMPMOTIFR
LFLAFMGMSIMNSAEGPAPQVAVSLAACVYELNEKTKSGRACLLDIGALVAFMF
CGSLTPMPTFLIOPTWLTBELTSLVAYVFLSCTFK"
complement(join(21486..21596,21673..21885,21982..22080,
22156..22281,22380..22596,22864..22958,23171..23391,
23446..23549,23686..23811,23899..23939,24271..24475,
24564..24820,24909..25038,25122..25183))
/note="gene_id:MYJ24.4"
/codon_start=1
/evidence=not_experimental
/product="acetyl-CoA synthetase-like protein"
/protein_id="BAB09822.1"
/db_xref="GI:9759363"
/translation="MMVYGCYKDFGPDLPAMIPDPKVASLTNNGKLEKRGKFLGN
YKNPVSSSSFOEFSVSNPEVYKTYLDELNITFSPPKICLEKDTSGDNGSKWLPB
ATLNPARNCLNGFKRRLDIYLRWDESSDDLPVMTWLLERLQVWMLAAVLSLG
LEESAIAVDMPMNVESVILYLAIVLAGHVVSIADFSREISRLKISKAALITFO
DVIIRGDKSIPLYRVDADAPLAIYVPAWGSSRKMLREKIDLSMNNPGLGIAAIFTO
EYVAAKPGAGATNIIIFSSGTGIGIPMKRRHGEPAKIPMTNISPLKSADAMCHDY
ORGDDVAMPNIGMMGMPWLIVYASLTNGACMCIYNSSPLGPTFAKYODAEVSLYI
PSIVRTWONSNTSGYDMSRIRCFSGTSEANSIDETLMLGMRHNTFVLEYCGTEIG
GSFTISGLDPOSLLAASSTAMCKLFLILDESNPIPPYAGVGLALCPHMGASST
LNGNHFYVFGQMPFOQILRRHGDLEERTSKGYRAHGRADDTMNLGKIVGSTE
IERVNSVDSDVLETAIGVPPSGPEQLVIAVFKSEFSPNDLNLKKSNSSEO
KMLNPLEKSVVLTPLSPRTATNMYRRLROOLTOTGINSKL"
complement(join(26733..26967,27047..27178,27258..27399,
27550..27742,27842..28087,28170..28385))
/note="emb|CAB75797.1"
/gene_id="MYJ24.5"
similar to unknown protein"
/codon_start=1

CDS
/evidence=not_experimental
/protein_id="BAB09823.1"
/db_xref="GI:9759364"
/translation="MAAEATKSSLSAKLTLTPSSSTKKTLSLRQVSVSLPSTISL
LSLFASPPEHAKAAVSIIPKQIVSLTEVEKTIINOVEGSSVFDATORFOVGDAL
KPAIDTALPIAKQAGEAMKIASPASEKQMOGSGFSEEPVNAKAVTIVA
OOTSKAIEDAKPIASSTMDTISADPSVYVAAAGAVLILPVPFASISPFRRYK
GDLTPAOTIDLCITKNYLMWDIRSEKDEKACIPRLPSNAKNVTSIPLPEELPNKKG
IYANSKRVEEIALAKISYLKTKINKSNITIIIDSYDSAKIYAKTILVGLKNCYIVT
DGSPSGRWLQSDLSGTDSYNSFAQVLSPRITPAASRSFGTSGTKFLPSSD"
31967..32800
/note="contains similarity to thymidine kinase
/gene_id:MYJ24.6"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB09824.1"
/db_xref="GI:9759365"
/translation="MRTLISPLSLAPSLHLHKPLSFSTALRFSFINNITPTNSPST
ISIRKQTKATRYTSSSSQPLSSSSQPLSHVYVGMFGSKTTLRLRIARETKR
IATIKNKDTRVCTESIVTHDGEKYPQMSLPOLSEKFRGPDYENRLDIVGIDBAO
FRGDLTFECREADKRGKTYIVAGLDGDRRRRRGVLDIIPADIVYTKITSCEVCG
KRALFTMRKTEKETELIGAEVYMPVCSHYVCGQNVLETARAVLDDSSNNHNSVASS
L"
complement(join(33199..33828,33917..34099,34459..34587,
34667..34948,35075..35564,35726..35870,36157..36528,
36577..36652,36881..37020,37165..37233,37356..37417,
37498..37601,37818..37912,37989..38079,38728..38862))
/note="gene_id:MYJ24.7"
unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB09825.1"
/db_xref="GI:9759366"
/translation="MGSDEDEYVHGTPIEREELIARRKKAAYAGASGMLRTLPANKQ
EYVDEGRRRFRHAGFTGYSGAGYVNTVSGHEGAPQSFSSRVRNARGAKOSTLDFLD
EDBKADMEKSLSSASQPDFTGFTAAHSHAKKOEHPSPAIPEPVDELYAIPASE
SICADKLKMGMRGHSIKVIRASSDAPRRAPKAPLAFTDENTKFTPLSVETVE
TSLGEDIKISESPVYVLPKODHLGLDIPKHADEPFKGIAPAGSICALAEFDYED
EDYATGYDDQTYVIDEOPAROSDNRLRLTSKENDVLPGFCAKNSDSMERFPP
IIPKDVARRHKEFGPLEAFKPTVSAPEVPPADNNLKLLEGATFVSRCKLYED
RSLNEKSNQLEDFLEKNGHDYARLRLEDOQRKDQSLTLDVAVSPVOMKMET
RGSLKEQSLDRLKLETDIASGSGSFQPTNLSDFTFSASSQEAADAVKPKDPA
KROPEDFLKEKKGGLRTDSDNRVNSMSAARLDPEAAEAIEKGAKVEYR
ATOPRDLPLAGLOPFGSGTEOKDQGVDMKSSKPYPRFEQMPSPILCKRRPLD
DPFGKLPAPARRKMKDSIVLPLDTYKKAASAOVSESVPKKETSIEBEVEYVEN
VERPDLVYKAFESDSEDDQPMNGKIDEGQKKAALATTLNRIADDFLESLEKE
LLEGEVMEBEIKSRSPEDSDKRLDRPLGKKEVEKETSITLGSSEERSKRRKREKSP
GKSGGNDLSSSGSDERRRRKRYNKKDRHRDSESDSDYSDRQGSRSKRRRE
SSREKRSKHKSHKRRTKKSSSSRSRSDDEOKESRREKKRRRD"
join(39364..39418,39573..39662,39749..39912,40006..40176)
/note="gene_id:MYJ24.8"
sp|P49592"
/codon_start=1
/evidence=not_experimental
/product="TATA-binding protein-associated phosphoprotein
Dtl protein homolog"
/protein_id="BAB09826.1"
/db_xref="GI:9759367"
/translation="MDPMDIVGSKEDASLPKATWTKIIEMLPPDVYARDAODLLI
ECQEVFINVNSNDYCNKEDRTIAPHEVILALQVLGGEVTEEVYVAYEDHKET
MOTQOSVWNPSPADQMTFEEBAAEQRMFAEARANGVSVQPRHPHETDQSPGS"
43330..44363
/note="gene_id:MYJ24.9"
p|P1700822"
similar to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB09827.1"
/db_xref="GI:9759368"
/translation="MAKDFAASRHNFSFWTRKYGSDENDDVSSHKPLPHHNTKPSSS
SSSSSNVITKKKLOSFAVSRLSVIAITLSRAPNONSGLSRVVGTLFGSRRHV
HFSIQDPNSPAPFLIELATPIGLVEMASGLVRLALREDCGKEEEDGKGGTLDH
GGGDKTKTTTAAVSRRLVEEPMWRTYCGKKGCGFATLRDCGKKEKKVILKALEMVGSG
```


VERSION AC069017.20 GI:15559167
KEYWORDS HTGS_PHASE1: HTGS_DRAFT: HTGS_FULLTOP.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 174140)
AUTHORS Metzger M.L., Lewis L.R., Hume J., Edwards C., Harris C.,
Dederich D., Thomas S., Okunou G., Carlock C., Garner T.,
Addison S., Pace A., Williams G., Bonnin D., Brooks A., Brown J.,
Buhay C., Bunac C., Burkett C., Chacko J., Chen G., Chen Z.,
Cox C., Davis C., Delgado O., Ding Y., Dugan-Rocha S.,
Fernandez C., Ferraguto D., Forcum-Tansey J., Gill R.,
Gorell J.H., Gunaratne P., Haller G., Hernandez J., Hognes M.,
Hosak H., Hou X., Huber J., Jackson L., Jia X., Kelly J., Kelly S.,
Kovat C., Liu J., Liu W., Louised H., Lozada R.J., Martin R.,
Massey E., McLeod M.P., Mei G., Moore S., Morgan M., Morris S.,
Neal D., Nelson A., Nguyen R., Nguyen N., Ogih M., Parish B.,
Perez L., Reiter D., Say J., Shen H., Vasquez L., Watlington S.,
Williamson A., Wrenford G., Zhou X., Bouck J., Hodgson A.,
Muzny D.M., Rives M., Scherer S., Sodergren E., Weinstein G.,
Worley K. and Gibbs R.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 174140)
AUTHORS Morley K.C.
TITLE Direct Submission
JOURNAL Submitted (17-MAY-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
COMMENT On Sep 12, 2001 this sequence version replaced gi:14787161.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: MMS3-342116
Center clone name: MMS3-342116
----- Summary Statistics
Sequencing vector: M13: L08821
Chemistry: Dye-primer BodyPy: 37% of reads
Chemistry: Dye-terminator Big Dye: 63% of reads
Assembly program: Phrap: version 0.990329
Consensus quality: 175325 bases at least Q40
Consensus quality: 179254 bases at least Q30
Consensus quality: 181014 bases at least Q20
Estimated insert size: 178322: sum-of-coverage
Quality coverage: 0x in Q20 bases; sum-of-coverage
Quality coverage: 7.4x in Q20 bases; sum-of-coverage
----- NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 86925: contig of 86925 bp in length
* 86926 87025: gap of unknown length
* 87026 87025: contig of 49703 bp in length
* 136729 136828: gap of unknown length
* 136829 159942: contig of 23114 bp in length
* 159943 160042: gap of unknown length
* 160043 168814: contig of 8772 bp in length
* 168815 168914: gap of unknown length
* 168915 174140: contig of 5226 bp in length.
* Location/Qualifiers
1. 174140
/organism="Mus musculus"

/db_xref="taxon:10090"
BASE COUNT 51417 a 32907 c 34252 g 55160 t 404 others
ORIGIN
Query Match 4.1%; Score 58; DB 2; Length 174140;
Best Local Similarity 56.2%; Pred. No. 0.0014; Mismatches 85; Indels 0; Gaps 0;
Matches 109; Conservative 0; Mismatches 85; Indels 0; Gaps 0;
QY 235 catccagatgatagatattgatgatagatagatgagcccatgcaaggccagcattgt 294
DB 42084 CATTCGATGACACCCCTGTTAACTAAATGCTTGATTTCAACACCGCCTCTAT 42143
QY 295 gtgtacacacacccagcatttaccagaaatgtagtaggaatcgaatcttgaaa 354
DB 42144 TTGTAAACACATGCTGTTTACCCGCCAGAGACATTCATGCAATGAGAAATGCAAGA 42203
QY 355 ggcacgaagaagggaatcttataaactgacagatggaatagatcattctgtg 414
DB 42204 AGCAGGAAAAAAGATGATCCCTGAACTTTGGAGGTTTGAAATGAGTTGATCTGTC 42263
QY 415 tatcatcacaga 428
DB 42264 TACCATCTAACAG 42277
RESULT 10
PFMAL13P3/C 318221 bp DNA linear HTG 19-AUG-1999
LOCUS Plasmidium falciparum chromosome 13 strain 3D7, *** SEQUENCING IN
DEFINITION PROGRESS ***, in unordered pieces.
ACCESSION AL049184.5 GI:5763803
VERSION AL049184.5
KEYWORDS HTG: HTGS_PHASE1.
SOURCE malaria parasite P. falciparum.
ORGANISM Plasmidium falciparum
REFERENCE 1 (bases 1 to 318221)
AUTHORS Bowman S., Churcher C., Harris B., Harris D., Lawson D., Quail M.
and Barrett J.B.
TITLE Direct Submission
JOURNAL Submitted (15-MAR-1999) P. falciparum Genome Sequencing Consortium,
The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
CB10 1SA, UK
COMMENT On Aug 24, 1999 this sequence version replaced gi:5731882.
For more information about this sequence or the Malaria Project,
see http://www.sanger.ac.uk/projects/P-falciparum. IMPORTANT: This
sequence is unfinished and does not necessarily represent the
correct sequence. Work on the sequence is in progress and the
release of this data is based on the understanding that the
sequence may change as work continues. The sequence may be
contaminated with foreign sequence from E.coli, yeast, vector,
phage etc.
Order of segments is not known; 800 n's separate segments.
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1. 318221
Location/Qualifiers
/organism="Plasmidium falciparum"
/strain="3D7"
/db_xref="taxon:5833"
/chromosome="13"
BASE COUNT 113610 a 29264 c 27321 g 118405 t 29621 others
ORIGIN
Query Match 3.9%; Score 54.4; DB 2; Length 318221;
Best Local Similarity 49.1%; Pred. No. 0.0092;
Matches 172; Conservative 0; Mismatches 176; Indels 2; Gaps 1;

RESULT 11			
AX083744			
LOCUS			
DEFINITION	AX083744	1141 bp	DNA
	Sequence 22 from Patent WO0111061.		Linear
			PAT 28-FEB-2001

	/note="consensus sequence of A.t., L.a., and B.n. FAEI promoters"			
BASE COUNT	123 a	32 c	42 g	112 t
ORIGIN	832 others			

Query Match	3.7%	Score 51.8;	DB 6;	Length 1141;
Best Local Similarity	9.4%	Pred NO. 0.083;		
Matches 75;	Conservative 318;	Mismatches 397;	Indels 6;	Gaps 2;

[illegible]

Db 311 SKTKRVPTSCVANNCCBAGDANKDHKKWKKSAAMGVYNNNNNNNNMMWTYKKARHBAEWDVW 370

Qy 893 ctaaaagaggcagctcaagaacataccagttcaacaataaactatactatgtatgactg 952

Db 371 HSAMKWHANAHYSRKKWTBYKRTYVNNNGTTTMMRMAVYWKMDMBGTNNNN 430

Qy 953 aggaactcgaagaaattctactacgttgctgaatttgaatagatcagcgtttcaagta 1012

Db 431 GGRITYGTTKKKKMWTYTKWKANCKKWRAMDHKTHTHNTTWKMKTYWNNCKYMSKSTNG 490

Qy 1013 tggagaagaatctctaaagt--gtcatacagctccacaagaacccaagatactactctt 1069

Db 491 KSHBBAALVYTWYMMWMMRRYAHANNNNNDYMKKACTWKKYEVCSKMMNNYAAWYKSSMN 550

Qy 1070 tcccaagctgtgtgagtagctgcctgcgtacattactacacaataaaaaaccatagagcct 1129

Db 551 YTSRYEHWKTNNSSWFRSDTRSMGRANNYARABHGYKWNTRMWBESHTEWBHBRAGAAY 610

Qy 1130 tctgtttttgcctctctctcttggaactggcgctgcattccatcagtgtaatggccacttg 1189

Db 611 WMBMWYBAKCKCMKAWYAKRYAGGSGNNNNNNNNNNNNNNNNNNATCARADYIASRYA 670

Qy 1190 cactgagatccagcagaagacacccgtgtgcgtgatgaatcagagtgctgttcgaagt 1249

Db 671 MA---NAKWTYKKBANNAAYTTHANNWGCMMNNADTDRITMKNNNNNNNAGTGWKNNNNN 727

Qy 1250 actgaataacagtttaatgaacagcatataatagctcctgcatalatgtaatgtaataca 1309

Db 728 AKNSAAKNVAAALAAKAKKHWRMANKWAMGWHADAABTTDKRNCAGATYKTYTNNNN 787

Qy 1310 gltaaaaaacggtatctccctgctgtctgatccacaacatcagtgtaacagaacacccc 1369

Db 788 TYRCVVTATTAARDGANN 847

Qy 1370 tatcagtggtgaaag 1385

Db 848 WINKWYTTTDDRWB 863

RESULT	12
AF151112	
LOCUS	2389 bp mRNA Linear INV 29-MAR-2000
DEFINITION	Dictyostelium discoideum cheater ChnA (chn) mRNA, chn-A allele,
ACCESSION	AF151112
VERSION	partial cds.
KEYWORDS	AF151112.1 GI:5007063
SOURCE	.
ORGANISM	Dictyostelium discoideum. Dictyostelium discoideum Eukaryota; Mycelozoa; Dictyostellida; Dictyostellium.
REFERENCE	1 (bases 1 to 2389) Emnis,H.L., Dao,D.N., Pukatzki,S.U. and Kessin,R.H. Dictyostelium amoeba lacking an F-box protein form spores rather than stalk in chimeras with wild type Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3292-3297 (2000)
TITLE	20202628 2 (bases 1 to 2389) Dao,D., Emnis,H.L., Franke,J., Kessin,R.H., Nelson,M.K. and Pukatzki,S.. A F-box protein is essential for development in Dictyostelium discoidium unpublished 3 (bases 1 to 2389) Dao,D., Emnis,H.L., Franke,J., Kessin,R.H., Nelson,M.K. and Pukatzki,S. Direct Submission Submitted (13-May-1999) Anatomy and Cell Biology, Columbia University, 630 W 168 Street, New York, NY 10032, USA Location/Qualifiers 1..2389 /organism="Dictyostelium discoideum" /strain="AX3K" /db_xref="taxon:44689"
JOURNAL	
FEATURES	
SOURCE	


```

RESULT 14
AF151111 3985 bp DNA linear INV 29-MAR-2000
LOCUS Dictyostellium discoideum cheater Chta (cht) gene, cht-A allele,
DEFINITION complete cds.
ACCESSION AF151111
VERSION AF151111
KEYWORDS AF151111.1 GI:5007061
SOURCE Dictyostellium discoideum.
ORGANISM Dictyostellium discoideum.
REFERENCE 1 (bases 1 to 3985)
AUTHORS Emsls,H.L., Dao,D.N., Pukatzki,S.U. and Kessin,R.H.
TITLE Dictyostellium amoebae lacking an F-box protein form spores rather
than stalk in chimeras with wild type
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3292-3297 (2000)
MEDLINE 20202628
MEDLINE 2 (bases 1 to 3985)
AUTHORS Dao,D., Emsls,H.L., Franke,J., Kessin,R.H., Nelson,M.K. and
Pukatzki,S.
TITLE A F-box protein is essential for development in Dictyostellium
discoideum
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 3985)
AUTHORS Dao,D., Emsls,H.L., Franke,J., Kessin,R.H., Nelson,M.K. and
Pukatzki,S.
TITLE Direct Submission
JOURNAL Submitted (13-MAY-1999) Anatomy and Cell Biology, Columbia
University, 630 W 168 Street, New York, NY 10032, USA
FEATURES
source
location/Qualifiers
1..3985
/organism="Dictyostellium discoideum"
/strain="AX3K"
/db_xref="taxon:44689"
/chromosome="2"
join(1..237,379..3885)
/gene="cht"
/product="cheater Chta"
<1..3885
/gene="cht"
/allele="A"
join(1..237,379..3885)
/gene="cht"
/note="F-box protein with WD40 domains"
/codon_start=1
/product="cheater Chta"
/protein_id="AAD37799.1"
/db_xref="GI:5007062"
/translation="MOYVNGNDISSQSIHEVLEAFSTKIPWTKLEESSEKHIHSY
SSHERGIYLSQVAEENISPEKNMYELDISRSKMDPFCRAHITIEEDHNIHL
NETNPILNDMLNLRSCYDQDFRFLVAMRSIELEDQDVOFELPAGWVLOIGRG
QKDKCKMPVQCCDLADIELQRIPIKFSNKRERLEDFULFPAVYSRAKIFES
IELYISNNVADLETKDIRISTEKAKEVNEFGTNPYGMKITYKIKJDMELLIKT
TSGYIMIGGSFSRYSPELLADVLVYQKNPEMDYFDKLVESINSSIREVEVHYR
MARNITIMRLDLSVKKGPONFSSVHRSICSPNYGADIEVHYPTALNGLDGS
FSPFLAIEIEKGYPTMEBEENVTKMEAFRIINSONSIINHVKMGKSMRELPIV
SVYHCGDVGSHIGTPYVNAVENPADILDEGNKKGWVSSKRNKRNKNKNKN
NNDNIDEIIEBEENIEKDOOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO
NHSQLEKIKTPRESIRFTYLSNSSSGRPVIAVESKOPFLFLVSDSEKRGKOR
KTYVNSNGEDNLEPEVVOIITSNLSAINIVMLSVCKRFKATDSPIPLMKLYKNP
LEHKTPKRRQIHSNLSNDENKVNNSGDSADGSSQDEEQOQOQOQOQOQOQOQOQO
QOQOQEDLSNINLNDPISLWFTNGIIPONLSPTIEIIWVGIVSOLIDNDLSNETL
IOQLPHOVSLIOMNLDLAKIOHLSMISGMPATIGOESPINKNSDNPKPNKRNKN
NSYIDEEFINKMSHYTEKHOSKRWVNEPRITRPLKHNRAIKAYKEGSAITVS
TEKIKFWMLNTGOCIGDEGESVLSVEYDHTOKSSCIIWPLSDTKVHIGKNGVT
AVDFLEQPLEVHTSRPNVLADGDFTEPGKYLIEWHTIHLWDETSLIWKNEVIAH
TKKITOSKIVAQOHELKNGKIYFTSSDSAKAWDLNPGICINTLIVGSAVNCIEB
DYMALGSDTKLKLMDLRQASTPISSYSYKTRGPRCISYOEKNGIYSGSDGSI
AFNLDNMNLSNLSVKKPIFNNVNIGNICITTTTTTTTTTTTTTTTTTTTTTTT
INTGPKDSKRLHNHESAVTCIESDEBAGTISGSQNGVLKMPD"
BASE COUNT 1559 a 575 c 608 g 1243 t
ORIGIN

```

```

Query Match 3.2% Score 45; DB 3; Length 3985;
Best Local Similarity 49.0%; Pred. No. 2.7; Indels 0; Gaps 0;
Matches 120; Conservative 0; Mismatches 125;
QY 694 gtccagaatctttggacaacatgcgtccagatggggcagaactcttaattcttaaat 753
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1495 gtccacacatcatgttggtgatgttggttcattatgtgactcctaagctaatatgca 1554
QY 754 cacatggaaaaaatcttatttctgaatagataagaagatctgagctcctaattgctg 813
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1555 atgctgcagaaattttgctgtatgactttttgaaagagatgaaatgaaatgaaatg 1614
QY 814 taticagtaaggcgaacaatcacagatggagacagattgaaagaagaacattcatgca 873
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1615 tcttcaaaaaattaaagcaaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatg 1674
QY 874 tctgtaattgatagttactaaagaagcagctcaaaagacataccagttcaacaata 933
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1675 aatataattatgatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatg 1734
QY 934 accta 938
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1735 accca 1739
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 15
HS425C14 160203 bp DNA linear PRI 23-NOV-1999
LOCUS HS425C14/c
DEFINITION Human DNA sequence from clone 425C14 on chromosome 6q22 Contains
the HSF2 gene for Heat Shock Factor 2 (Heat Shock Transcription
Factor 2, HSF 2) and an unknown gene similar to the placental
protein DIFP3 gene. Contains ESTs, STRs and GSSs, complete
sequence.
ACCESSION 299129
VERSION 299129.1 GI:3281967
KEYWORDS HTG; DIFP3; Heat Shock Transcription Factor; HSF2; HSF2.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 160203)
AUTHORS Mashreqhi-Mohammadi,M.
TITLE Direct Submission
JOURNAL Submitted (03-JUL-1998) E-mail enquiries: humquerry@sanger.ac.uk
COMMENT Clone requests: clonerequest@sanger.ac.uk
On Jul 1, 1998 this sequence version replaced gi:3250834.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above. This sequence is the
entire insert of clone 425C14. This sequence has been finished
according to sequence map criteria as follows. An attempt is made
to resolve all sequencing problems, such as compressions and
repeats, but not necessarily within known annotated human repeat
sequence elements (e.g. Alu). Where the sequence is ambiguous,
there is an annotation using the "unsure" feature key.
This sequence was generated from part of bacterial clone contigs of
human chromosome 6, constructed by the Sanger Centre Chromosome 6
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr6
425C14 is from the library RPI3 constructed at the Roswell Park
Cancer Institute by the group of Pieter de Jong. For further
details see http://bacpac.med.buffalo.edu/VECTOR:pcypac2.
location/Qualifiers
1..160203
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="q22"

```

```

repeat_region 22624..22769
/feature="MIR repeat: matches 49..199 of consensus"
repeat_region complement(26830..27596)
/feature="LI repeat: matches 4031..3265 of consensus"
misc_feature 29432..29832
/feature="HSF2"
misc_feature /note="match: STS G13562"
misc_feature complement(29857..30169)
misc_feature /note="match: STS G05769"
polya_signal 30148..30153
/feature="HSF2"
polya_site 30171
/feature="HSF2"
polya_site /gene="HSF2"
polya_site /gene="HSF2"
repeat_region complement(31226..31515)
repeat_region /note="AluX repeat: matches 292..2 of consensus"
repeat_region complement(32572..35874)
repeat_region /note="AluX repeat: matches 301..1 of consensus"
repeat_region 32902..33204
/feature="AluY repeat: matches 1..299 of consensus"
misc_feature 34025..34261
/feature="match: STS G24632"
repeat_region 34383..34412
/feature="15 copies 2 mer tg 93% conserved"
repeat_region 35013..35126
/feature="LI repeat: matches 2394..2506 of consensus"
repeat_region 35128..35538
/feature="MST repeat: matches 1..426 of consensus"
repeat_region 35539..35586
/feature="MST-INTERNAL repeat: matches 1..48 of consensus"
repeat_region 35578..36192
/feature="LI repeat: matches 1057..1737 of consensus"
repeat_region 35600..36060
/feature="MER25 repeat: matches 1583..2099 of consensus"
repeat_region 36219..37029
/feature="MST-INTERNAL repeat: matches 807..1651 of consensus"
repeat_region 37030..37447
/feature="MST repeat: matches 1..426 of consensus"
repeat_region 37541..38260
/feature="LI repeat: matches 3171..3958 of consensus"
repeat_region 38261..38558
/feature="AluX repeat: matches 1..302 of consensus"
repeat_region 38566..39641
/feature="LI repeat: matches 3943..5041 of consensus"
repeat_region complement(39724..40048)
repeat_region /note="L1M1 repeat: matches 749..390 of consensus"
repeat_region complement(join(40559..42390,43985..44215,44332..44476,44877..44896,45099..49268,50981..51118,51362..51441,55692..53661,55731..>55892))
/feature="dU25C14.2"
/feature="match: CDNAS U49188 L29441, match: ESTs R63311
AA642883 AA680098 AA757435 AA147658 AA364378 N23258 N73280
AA563775 AA479364 N68017 H96625 R64053 R62561 AA054470
N29921 AA587827 AA718993 AA099130 D55899 AA052380 R80437
AA661998 AA056751 R63339 AA946958 AA788657 N47022 AA855136
AA640774 AA359308 T77723 N26641 D60199 C14222 AA333766
AA645582 AA838780 W03704 AA058510 T05049 N75278 AA903139
R74572 C16158 AA573354 C16394 AA628814 AA706727 R41117
AA426675 H34768 A1003241 T34540 T10566 W030676 W39974
AA426614 W32165 C18045 W23860 AA792501 N31789 N87445
AA426825 AA238110 R10545 R57569 AA099167 H96624 H47943
W76583 AA298096 AA825649 Z42747 T32615 H47954 N31858
H70513 AA010230 W02277 AA178232 AA2296204 AA662364 AA178351
R62199 AA318861 Z44998 D60904 D52663 R12298 AA137672
R11667 R82588 H35640 AA590123 W97491 AA663768 R18882
AA658854 AA657717 AA657439 AA000956 AA666314 AA087604
W52428 T59284 AA068123 AA517373 T59243 AA915293 AA636161
AA984271 AA982116 AA096521 AA852720 AA986636 AA853329
AA344622
/feature="dU25C14.2 (Placental protein D1FF33 LIKE)"
/feature="not experimental"

```

complete - "Airtub Rep
/note-

425C14.2"

40565. 40844

```
/note="match: STSS G05764 G22324 G20260  
complement(40576..40581)
```

`/gene="dJ425C14.2"`

Score 44.6; DB 9; Length 160203;

```

...:20;      Fred: NO. 2;
... 0; Mismatches 234; Indels 0; Gaps 0;

```

aaagaaaatttqaaaaagcttcttcaaaatatctgatatqca 195

AGGAAGAATAATCTCAATCCAGATTACTATGGGCTGTA CT 5865

[illegible]

ccgcgcccgaaccagacatccagctgcagaataattc 255

AAATGATTTCCTCAGATTATTATCAATATACAAAATT 5805

attgcaaggccagcacttgtgtgtacacaacacgacattt 315

CCAGCAACAGCTAGAATCTATTTTAAGTCAAGACACCATT 5745

gggaattcagaatcttggaaaagcgcgaagagggaaatcct 375

TTCCAGGAGAAATGAAACAAGATCCACCAAGGCTAATAGCA 5685

aaataaqaattcaattctatataatcataatcacaaactgccca 435

[illegible]

GGAAATTAGGAAATATGCTTGTGCAATGAGAAATACATCAATCAGG 3023

gacatccctgtgtatcttctgataccctcatgccaagatatgccca 495

ATCTGATTATATAGTTTGGATCCAATTGCAGCAAAAATCCAA 5565

cccgacgcacatgttttagagatttgatgcagattttaggac 554

ATGGACCTTGATTTTAAACATATGTGGAAGAGCTAAAGAC 506

06:53:41

```
Search completed: May 22, 2002, 06:53:41
Job time: 8806 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 22, 2002, 06:43:32 ; Search time 463.88 Seconds
(without alignments)
5185.387 Million cell updates/sec

Title: US-09-693-205-7_COPY_5300_6700

Sequence: 1 gtacgcagtaaacatagagc.....aaagacacttaagaagt 1401

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

N_Geneseq_032802.*
1: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
22: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
23: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1401	100.0	12792	22	AAH20176	Human mutated spas
2	1390	99.2	12793	22	AAH20174	Human spastin nucl
3	1390	99.2	12793	22	AAH20178	Human mutated spas
4	1390	99.2	12793	22	AAH20182	Human mutated spas
5	1388.4	99.1	12793	22	AAH20179	Human mutated spas
6	1174	83.8	11493	22	AAH20175	Human spastin nucl
7	40.6	2.9	7990	24	ABU32158	Mouse immune syste
8	39.6	2.8	531	22	ABA59780	Human foetal liver
9	39.6	2.8	531	22	ABA28269	Probe #6735 for ge

C	10	39.6	2.8	531	22	AAK08051	Human brain expres
C	11	39.6	2.8	531	22	AAK33925	Human bone marrow
C	12	39.6	2.8	531	22	AAI16573	Probe #6506 for ge
C	13	39.6	2.8	531	22	AAI39650	Probe #8336 used t
C	14	39.6	2.8	10119	22	AAK87559	Human immune/haema
C	15	39.6	2.8	12247	22	AAK73076	Human immune/haema
C	16	39.6	2.8	12247	22	AAK87543	Human immune/haema
C	17	39.6	2.8	14983	22	AAK73075	Human immune/haema
C	18	39.6	2.8	14983	22	AAK87542	Human immune/haema
C	19	39.6	2.8	20991	22	AAK87545	Human immune/haema
C	20	39.6	2.8	32190	22	AAI62927	Human immune/haema
C	21	39.6	2.8	32249	22	AAI62932	Human genomic DNA
C	22	39.6	2.8	39068	22	AAK71820	Human immune/haema
C	23	39.6	2.8	39068	22	AAK73078	Human immune/haema
C	24	39.6	2.8	39068	22	AAK85294	Human immune/haema
C	25	39.6	2.8	39068	22	AAK87544	Human immune/haema
C	26	39.6	2.8	39110	22	AAK71825	Human immune/haema
C	27	39.6	2.8	39110	22	AAK73087	Human immune/haema
C	28	39.6	2.8	39110	22	AAK87555	Human immune/haema
C	29	39.6	2.8	45300	22	AAK73079	Human immune/haema
C	30	39.6	2.8	45300	22	AAK87547	Human immune/haema
C	31	39.4	2.8	815	21	AAF21681	Human breast and o
C	32	39	2.8	7503	21	AAAF0206	Plasmodium falcipla
C	33	38.6	2.8	1092	23	AA555809	Streptococcus pneu
C	34	38	2.7	21358	22	AA539919	Genomic sequence #
C	35	38	2.7	21358	22	AAI06419	Human reproductive
C	36	38	2.7	21358	22	AAK73090	Human immune/haema
C	37	38	2.7	21358	22	AAK87446	Human immune/haema
C	38	38	2.7	21358	22	AAK87558	Human immune/haema
C	39	38	2.7	21358	22	AAK90363	Human digestive sy
C	40	38	2.7	21676	22	AA539918	Genomic sequence #
C	41	38	2.7	21676	22	AAI06418	Human reproductive
C	42	38	2.7	21676	22	AAK73081	Human immune/haema
C	43	38	2.7	21676	22	AAK87445	Human immune/haema
C	44	38	2.7	21676	22	AAK87549	Human immune/haema
C	45	38	2.7	21676	22	AAK90362	Human digestive sy

ALIGNMENTS

RESULT	1
ID	AAH20176 standard; DNA: 12792 BP.
AC	AAH20176;
XX	
DT	09-AUG-2001 (first entry)
XX	
XX	Human mutated spastin nucleotide sequence SEQ ID NO:7.
XX	
KW	Human; mouse; spastin; ARSACS; chromosome 13q11; identification;
KW	autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;
KW	neurodegenerative disease; reduced sensory nerve conduction; diagnosis;
KW	reduced motor nerve velocity; hypermyelination of retinal nerve fibre;
KW	atrophy of upper cerebellar vermis; absence of Purkinje cell;
KW	abnormal neuronal lipid storage; genetic disorder; characterisation: ds.
XX	
OS	Homo sapiens.
OS	Synthetic.
XX	
FT	Key
FT	CDS
FT	Location/Qualifiers
FT	/*tag= a
FT	/product= "mutated spastin"
XX	
XX	MO200129266-A2.
XX	
PD	26-APR-2001.
XX	
PF	20-OCT-2000; 2000WO-US29130.
XX	
PR	20-OCT-1999; 99US-0160588.

XX (UYMC-) UNIV MCGILL.
PA (HOPI-) HOPITAL SAINTE-JUSTINE.
XX
XX
PI Hudson TJ, Engert J, Richter A;
XX
DR WPI: 2001-308494/32.
DR P-PSDB; AAB97821.
XX
XX New isolated polynucleotide, encoding spastin gene, and polypeptides,
PT useful for diagnosing autosomal recessive spastic ataxia of
PT Charlevoix-Saguenay disease by detecting two point mutations in spastin
PT gene sequence -
XX
PS Claim 1; Page -: 76pp; English.

CC The present invention describes human and mouse spastin, and mutated
CC human spastin (autosomal recessive spastic ataxia of Charlevoix-Saguenay
CC (ARSAACS)) gene sequences (1). The spastin gene has been mapped to
CC chromosome 13q11. (1) have neuroprotective activities and can be used in
CC gene therapy and as a spastin polypeptide agonists. (1), their fragments
CC or their complements can be useful for assaying the presence of a nucleic
CC acid molecule in a sample. (1) is useful for diagnosing or aiding in the
CC diagnosis of an early onset neurodegenerative disease in an individual.
CC The neurodegenerative disease comprises reduced sensory nerve conduction,
CC reduced motor nerve velocity, hypermyelination of retinal nerve fibres,
CC atrophy of upper cerebellar vermis, absence of Purkinje cells and
CC abnormal neuronal lipid storage. (1) can also be used to produce
CC antisense nucleic acids, is useful as molecular weight or chromosome
CC markers, to identify genetic disorders, as hybridisation probes or
CC primers, as an antigen, identify and express recombinant protein for
CC analysis, characterisation or therapeutic use, or as markers for tissues
CC in which the corresponding protein is expressed. Diagnostic methods from
CC the present invention can be used to identify subjects having or at risk
CC of developing a disease or disorder associated with aberrant expression
CC or activity of (1). The assays can be utilised to identify a subject
CC having or at risk of developing a disorder associated with Spastin
CC protein or spastin gene expression or activity. The present sequence
CC encodes a mutated human spastin.
CC N.B. The present sequence is not given in the present specification but
CC is derived from the human spastin nucleotide sequence (AAH20174) as
CC stated on page 14.

Sequence 12792 BP; 4163 A; 2256 C; 2487 G; 3886 T; 0 other;

Query Match	100.0%;	Score 1401;	DB 22;	Length 12792;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1401; Conservative	0;	Mismatches	0;	Gaps 0;

OY	1	gttgcagtaaaacttagagacagctcccaagcgacacaaagccttagaagaatatgcatcc	60
Db	5300	gttagcagtaaaacttagagacagctcccaagcgacacaaagccttagaagaatatgcatcc	5359
OY	61	aatgtctgttttacaaacacttggcacagaaatttgggcagagaagaaatattgaccagaga	120
Db	5360	aatgtctgttttacaaacacttggcacagaaatttgggcagagaagaaatattgaccagaga	5419
OY	121	attaagagcatccttaataatcatctcttcgaaagaagaaatgttgaagaagctcttcaa	180
Db	5420	attaagagcatccttaataatcatctcttcgaaagaagaaatgttgaagaagctcttcaa	5479
OY	181	aatgtctgaatgacaaagcgacagaaatctgttttgggttgatccttagacagcatcaa	240
Db	5480	aatgtctgaatgacaaagcgacagaaatctgttttgggttgatccttagacagcatcaa	5539
OY	241	gtttgatagaatatattgataagatgggcccattgacaagggccagcacttgatggttac	300
Db	5540	gtttgatagaatatattgataagatgggcccattgacaagggccagcacttgatggttac	5599
OY	301	aacaaacagcactttacagaagatgtgtttaagggaattcaagaatcttggaaaaagcagc	360
Db	5600	aacaaacagcactttacagaagatgtgtttaagggaattcaagaatcttggaaaaagcagc	5659

[illegible]

QY	1381	taaagcactttaagaagt	1401
Db	6680	taagcactttaagaagt	6700

RESULT 2
AAH20174
ID AAH20174 standard; DNA: 12793 BP.
XX
AC AAH20174;
XX
09-AUG-2001 (first entry)
XX
Human spastin nucleotide sequence SEQ ID NO:1.
XX
Human: mouse; spastin; ARSACS; chromosome 13q11; identification;
KW autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;
KW neurodegenerative disease; reduced sensory nerve conduction; diagnosis;
KW reduced motor nerve velocity; hypermyelination of retinal nerve fibre;
KW atrophy of upper cerebellar vermis; absence of Purkinje cell;
KW abnormal neuronal lipid storage; genetic disorder; characterisation; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 77..11566
FT /*tag= a
FT /product= "spastin"
XX
PN MO200129266-A2.
XX
26-APR-2001.
XX
20-OCT-2000; 2000MO-US29130.
XX
20-OCT-1999; 99US-0160588.
XX
20-OCT-1999; 99US-0160588.
XX
PA (UYMC-) UNIV MCGILL.
PA (HOP1-) HOPITAL SAINTE-JUSTINE.
XX
PI Hudson TJ, Engert J, Richter A;
XX
DR WPI; 2001-308494/32.
DR P-PSDB; AAB97819.
XX
XX
PT New isolated polynucleotide, encoding spastin gene, and polypeptides,
PT useful for diagnosing autosomal recessive spastic ataxia of
PT Charlevoix-Saguenay disease by detecting two point mutations in spastin
PT gene sequence -
XX
PS Claim 1; Fig 9; 76pp; English.
XX
The present invention describes human and mouse spastin, and mutated
CC human spastin (autosomal recessive spastic ataxia of Charlevoix-Saguenay
CC (ARSACS)) gene sequences (I). The spastin gene has been mapped to
CC chromosome 13q11. (I) have neuroprotective activities and can be used in
CC gene therapy and as a spastin polypeptide agonists. (I), their fragments
CC or their complements can be useful for assaying the presence of a nucleic
CC acid molecule in a sample. (I) is useful for diagnosing or aiding in the
CC diagnosis of an early onset neurodegenerative disease in an individual.
CC The neurodegenerative disease comprises reduced sensory nerve conduction,
CC reduced motor nerve velocity, hypermyelination of retinal nerve fibres,
CC atrophy of upper cerebellar vermis, absence of Purkinje cells and
CC abnormal neuronal lipid storage. (I) can also be used to produce
CC antisense nucleic acids, is useful as molecular weight or chromosome
CC markers, to identify genetic disorders, as hybridisation probes or
CC primers, as an antigen, identify and express recombinant protein for
CC analysis, characterisation or therapeutic use, or as markers for tissues
CC in which the corresponding protein is expressed. Diagnostic methods from
CC the present invention can be used to identify subjects having or at risk
CC of developing a disease or disorder associated with aberrant expression
CC or activity of (I). The assays can be utilised to identify a subject
CC having or at risk of developing a disorder associated with spastin
CC protein or spastin gene expression or activity. The present sequence
CC encodes human spastin as given in the present invention.
XX
Sequence 12793 BP: 4163 A; 2256 C; 2487 G; 3887 T; 0 other;

Query Match 99.2%; Score 1390; DB 22; Length 12793;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1401; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
OY 1 gtacagtaaaactgagagcagctcccaagcagacacaaagccttagaagatgcatcc 60
|||||
Db 5300 gtacagtaaaactgagagcagctcccaagcagacacaaagccttagaagatgcatcc 5359
OY 61 aatgctgttttacacacacttggcagagaatttggcgcaagaagaattgacccagaga 120
|||||
Db 5360 aatgctgttttacacacacttggcagagaatttggcgcaagaagaattgacccagaga 5419
OY 121 attaagagatccttaatgcatatcctctgaaaggaattgttaaagagcttctcaa 180
|||||
Db 5420 attaagagatccttaatgcatatcctctgaaaggaattgttaaagagcttctcaa 5479
OY 181 aatgctgtatgtgcaagcgacagagaatctgtttgtttgtatccttagacagatcca 240
|||||
Db 5480 aatgctgtatgtgcaagcgacagagaatctgtttgtttgtatccttagacagatcca 5539
OY 241 gtgatagaataatttgatgataagtgggcccattgcaaggcgacacttgtgttac 300
|||||
Db 5540 gtgatagaataatttgatgataagtgggcccattgcaaggcgacacttgtgttac 5599
OY 301 aacaaaccgcaatttacagaagaatgttttagaggaaatcagaatcttggaaaggcacg 360
|||||
Db 5600 aacaaaccgcaatttacagaagaatgttttagaggaaatcagaatcttggaaaggcacg 5659
OY 361 aaagagggaatccttataaaacttgcagagatagatagattcaattctgtatcat 420
|||||
Db 5660 aaagagggaatccttataaaacttgcagagatagatagattcaattctgtatcat 5719
OY 421 atcacagactgcccatcttcttcttcttgcgaatgacatccctgtatctttagatcccat 480
|||||
Db 5720 atcacagactgcccatcttcttcttcttgcgaatgacatccctgtatctttagatcccat 5779
OY 481 gccagatatgccacgaggggccacatcatatagctcccgagcgcatgtttagagatttgat 540
|||||
Db 5780 gccagatatgccacgaggggccacatcatatagctcccgagcgcatgtttagagatttgat 5839
OY 541 gcaagttttagagacacatttcagagatttccgagatttcttacttcttggaaacattttaa 600
|||||
Db 5840 gcaagttttagagacacatttcagagatttccgagatttcttacttcttggaaacattttaa 5899
OY 601 ctgataattgcacatgttccagatttcccttcttctgtaatgacagaattggcaaaagtctcg 660
|||||
Db 5900 ctgataattgcacatgttccagatttcccttcttctgtaatgacagaattggcaaaagtctcg 5959
OY 661 gaaattcgtctgttccagcatcagacagaatgtgtccagaatccttttgacaaactgtgc 720
|||||
Db 5960 gaaattcgtctgttccagcatcagacagaatgtgtccagaatccttttgacaaactgtgc 6019
OY 721 tcaagttggggcgaactcttaattgttctttaaaccatggaagaaatttcttctgtgaa 780
|||||
Db 6020 tcaagttggggcgaactcttaattgttctttaaaccatggaagaaatttcttctgtgaa 6079
OY 781 atagataagatgacgagctctaattgttcgttattcagtaaaaggcaaatccacagat 840
|||||
Db 6080 atagataagatgacgagctctaattgttcgttattcagtaaaaggcaaatccacagat 6139
OY 841 ggaagacagattgaaagaaacaatttcaatgcatctgttaattgtatgagtactataaag 900
|||||
Db 6140 ggaagacagattgaaagaaacaatttcaatgcatctgttaattgtatgagtactataaag 6199
OY 901 aggcagctcaagaagatcacagtttcaacaataaacttactatgatactgaagactct 960
|||||
Db 6200 aggcagctcaagaagatcacagtttcaacaataaacttactatgatactgaagactct 6259
OY 961 gaaggaatacttactactgtgctaaattgttaataatcagagcttccaagtaatggagaa 1020
|||||
Db 6260 gaaggaatacttactactgtgctaaattgttaataatcagagcttccaagtaatggagaa 6319

QY	1021	gtactctaaagtgctacatactacgtctcaagaacccaagatactactctttccacgtggt	1080
Db	6320	gtactctaaagtgctacatactacgtctcaagaacccaagatactactctttccacgtggt	6379
QY	1081	ggaatgacgtccgcgaacttactccaactctaaaaaccccaaggccttcgttttttg	1140
Db	6380	ggaatgacgtccgcgaacttactccaactctaaaaaccccaaggccttcgttttttg	6439
QY	1141	cctctctctcttggaactggcgccgcatctctcgtgaaatggccacttgcgactgattca	1200
Db	6440	cctctctctcttggaactggcgccacttcaatggaatggccacttgcgactgattca	6499
QY	1201	gccagaagaacctgtggcgtgatagtaaatgagtggtgttcgaagtgcagtgaataac	1260
Db	6500	gccagaagaacctgtggcgtgatagtaaatgagtggtgttcgaagtgcagtgaataac	6559
QY	1261	agtttaatgcaacatctaataagtcctctgatatg_tgatatgctatacaagttaaaaaa	1319
Db	6560	agtttaatgcaacatctaataagtcctctgatatg_tgatatgctatacaagttaaaaaa	6619
QY	1320	cggatctccctgttctgtgccaacatatcaagtgttaacgaacaccccatcattcagtt	1379
Db	6620	cggatctccctgttctgtgccaacatatcaagtgttaacgaacaccccatcattcagtt	6679
QY	1380	gtaaagacactttaagaagt_1401	
Db	6680	gtaaagacactttaagaagt_6701	

human spastin (autosomal recessive spastic ataxia of Charlevoix-Saguenay (ARSAOS)) gene sequences (1). The spastin gene has been mapped to chromosome 13q11. (1) have neuroprotective activities and can be used in gene therapy and as a spastin polypeptide agonists. (1), their fragments or their complements can be useful for assaying the presence of a nucleic acid molecule in a sample. (1) is useful for diagnosing or aiding in the diagnosis of an early onset neurodegenerative disease in an individual. The neurodegenerative disease comprises reduced sensory nerve conduction reduced motor nerve velocity, hypermyelination of retinal nerve fibres, atrophy of upper cerebellar vermis, absence of Purkinje cells and abnormal neuronal lipid storage. (1) can also be used to produce antisense nucleic acids, is useful as molecular weight or chromosome markers, to identify genetic disorders, as hybridisation probes or primers, as an antigen, identify and express recombinant protein for analysis, characterisation or therapeutic use, or as markers for tissues in which the corresponding protein is expressed. Diagnostic methods from the present invention can be used to identify subjects having or at risk of developing a disease or disorder associated with aberrant expression or activity of (1). The assays can be utilised to identify a subject having or at risk of developing a disorder associated with Spastin protein or spastin gene expression or activity. The present sequence CC represents a mutated human spastin gene from the present invention. CC N.B. The present sequence is not given in the present specification but CC is derived from the human spastin nucleotide sequence (AAH20174) as CC stated on page 14. CC XX

Sequence 12793 BP; 4163 A; 2255 C; 2487 G; 3688 T; 0 other:

Query Match	99.28;	Score 1390;	DB 22;	Length 12793;
Best Local Similarity	99.98;	Pred. No. 0;		
Matches 1401;	Conservative 0;	Mismatches 0;	Indels 1;	Gaps 1

RESULT 3
 ID AAH20178 standard; DNA; 12793 BP.
 XX
 AC
 XX AAH20178;
 DT 09-AUG-2001 (first entry)
 XX
 DE Human mutated spastin nucleotide sequence SEQ ID NO:11.
 XX
 KW Human; mouse; spastin; ARSACS; chromosome 13q11; identification;
 KW autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;
 KW neurodegenerative disease; reduced sensory nerve conduction; diagnosis;
 KW reduced motor nerve velocity; hypermyelination of retinal nerve fibre;
 KW atrophy of upper cerebellar vermis; absence of Purkinje cell;
 KW abnormal neuronal lipid storage; genetic disorder; characterisation; ds
 XX
 OS Homo sapiens.
 OS Synthetic.
 XX
 XX WO200129266-A2.
 XX
 XX 26-APR-2001.
 XX
 XX 20-OCT-2000; 2000WO-US29130.
 XX
 XX 20-OCT-1999; 99US-0160588.
 XX
 XX (UYMC-) UNIV MCGILL.
 XX PA (HOPI-) HOPITAL SAINTE-JUSTINE.
 XX
 XX Hudson TJ, Engert J, Richter A;
 XX
 XX WPI: 2001-308494/32.
 XX
 XX
 XX New isolated polynucleotide, encoding spastin gene, and polypeptides,
 PT useful for diagnosing autosomal recessive spastic ataxia of
 PT Charlevoix-Saguenay disease by detecting two point mutations in spastin
 PT gene sequence -
 XX
 XX Claim 1; Page -: 76pp; English.
 XX
 CC The present invention describes human and mouse spastin, and mutated

OY	1	gtaacagtaaaactaggagacagatgcccaagaagcgacacaaagcctttgaagaatatgacatcc	60
Db	5300	gtacgcagtaaaactaggagacatgcccaagaagcgacacaaagcctttgaagaatatgacatcc	5359
OY	61	aatctcgttttaacaacacttgcacagaatttggcgaaagaagaaatltgacgcgcaga	120
Db	5360	aatctcgttttaacaacacttgcacagaatttggcgaaagaagaaatltgacgcgcaga	5419
OY	121	ataagaagcatccttaattgacatacctctctgaaagaagaatgttgaaagacttctca	180
Db	5420	ataagaagcatccttaattgacatacctctctgaaagaagaatgttgaaagacttctca	5479
OY	181	aatgcgtatgatgtgaaagcgagcaagaatctgttttggttttgatccctagacagatcca	240
Db	5480	aatgcgtatgatgtgaaagcgagcaagaatctgttttggttttgatccctagacagatcca	5539
OY	241	gttcataagaatatttgaatgaatggggcccccctgcgaagggccgcgaacttgtgtgac	300
Db	5540	gttcataagaatatttgaatgaatggggcccccctgcgaagggccgcgaacttgtgtgac	5599
OY	301	aacaacagcccatltaagaagaatgaatgttagagaatcaagaatcttgaaaaagcgacg	360
Db	5600	aacaacagcccatltaagaagaatgaatgttagagaatcaagaatcttgaaaaagcgacg	5659
OY	361	aaagaagggaatccttctaacttgcacgctatggaatagatcctcaattcgtgtacat	420
Db	5660	aaagaagggaatccttctaacttgcacgctatggaatagatcctcaattcgtgtacat	5719
OY	421	atccacagctggcccatcttttatcttctgcgaatgaatatcctgtatcttggatcccat	480
Db	5720	atccacagctggcccatcttttatcttctgcgaatgaatatcctgtatcttggatcccat	5779
OY	481	gccagatatgcacacaggggcacatcccatltaglcccgagcgatgtttagagatttgat	540
Db	5780	gccagatatgcacacaggggcacatcccatltaglcccgagcgatgtttagagatttgat	5839
OY	541	gcgaattttgacacagttctcagatgtcttgabatttactctggaacccatttaa	600
Db	5840	gcgaattttgacacagttctcagatgtcttgabatttactctggaacccatttaa	5899

```

QY 601 ctggataattgcacaatggtcagattctcctcttcgtatgacgaagaatgacaaatggttcg 660
    |||
DB 5900 ctgataattgcacaatggttcagattctcctcttcgtatgacgaagaatgacaaatggttcg 5959
QY 661 gaatttcgtcttccagatcagaagaatggttcagaatcttctggacaactgcgc 720
    |||
DB 5960 gaatttcgtcttccagatcagaagaatggttcagaatcttctggacaactgcgc 6019
QY 721 tcagatgggacagaactctcaatggttcttaacacatggaataatcttatttgcga 780
    |||
DB 6020 tcagatgggacagaactctcaatggttcttaacacatggaataatcttatttgcga 6079
QY 781 atagataagtagtctggagctctaaatgctgtatcagtaaaaggacaaatcacagat 840
    |||
DB 6080 atagataagtagtctggagctctaaatgctgtatcagtaaaaggacaaatcacagat 6139
QY 841 ggaagacagattgaaagaagaacaaatctcatgcatctgtaattgtagttagtactaaag 900
    |||
DB 6140 ggaagacagattgaaagaagaacaaatctcatgcatctgtaattgtagttagtactaaag 6199
QY 901 aggcagctcaaaagacacacagttcaacaataaccatactatgatactgagactct 960
    |||
DB 6200 aggcagctcaaaagacacacagttcaacaataaccatactatgatactgagactct 6259
QY 961 gaaggaatcttactacgttgcataattgtcaatagatcagagcttctcaagatgagaa 1020
    |||
DB 6260 gaaggaatcttactacgttgcataattgtcaatagatcagagcttctcaagatgagaa 6319
QY 1021 gttctaaagatgcatatagctacagaagaacaaatatcttcttccacgtggt 1080
    |||
DB 6320 gttctaaagatgcatatagctacagaagaacaaatatcttcttccacgtggt 6379
QY 1081 ggggtgagctgcctgcattactacaaactataaaaccccatagagctctgtttttg 1140
    |||
DB 6380 ggggtgagctgcctgcattactacaaactataaaaccccatagagctctgtttttg 6439
QY 1141 cctcttcttctggagactggtgcctcatctcatgtgaatggccacttgcactgatatca 1200
    |||
DB 6440 cctcttcttctggagactggtgcctcatctcatgtgaatggccacttgcactgatatca 6499
QY 1201 gccagaaggaactgtggtgctgataatggaatggtgttcgaagtgcagtgaataac 1260
    |||
DB 6500 gccagaaggaactgtggtgctgataatggaatggtgttcgaagtgcagtgaataac 6559
QY 1261 agtttaatgacagcatatagctcctgcataatg-igaattgctcaataacgtttaaaaaa 1319
    |||
DB 6560 agtttaatgacagcatatagctcctgcataatg-igaattgctcaataacgtttaaaaaa 6619
QY 1320 cggatattcctctggtctgatacaacatctcaatggttacagaacacccattcatggt 1379
    |||
DB 6620 cggatattcctctggtctgatacaacatctcaatggttacagaacacccattcatggt 6679
QY 1380 gtaaaagacacttaagaagt 1401
    |||
DB 6680 gtaaaagacacttaagaagt 6701

```

RESULT 4

AAH20182 standard; DNA: 12793 BP.

AAH20182;

09-AUG-2001 (first entry)

Human mutated spastin nucleotide sequence SEQ ID NO:15.

Human; mouse: spastin; ARSACS: chromosome 13q11; identification:
 autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;
 neurodegenerative disease; reduced sensory nerve conduction; diagnosis;
 reduced motor nerve velocity; hypermyelination of retinal nerve fibres;
 atrophy of upper cerebellar vermis; absence of Purkinje cell;
 abnormal neuronal lipid storage; genetic disorder; characterisation; ds.

```

XX XX Homo sapiens.
OS Synthetic.
XX XX
FH Key Location/Qualifiers
FT CDS 77..1156
FT /tag= a
FT /product= "mutated spastin"
XX XX
PN MO200129266-A2.
XX XX
PD 26-APR-2001.
XX XX
PE 20-OCT-2000; 2000WO-US29130.
XX XX
PR 20-OCT-1999; 99US-0160588.
XX XX
PA (UYMC-) UNIV MCGILL.
PA (HOP1-) HOPITAL SAINT-Justine.
XX XX
PI Hudson TJ, Engert J, Richter A;
XX XX
DR WPI: 2001-308494/32.
DR P-PSDB: AAB97823.
XX XX
PT New isolated polynucleotide, encoding spastin gene, and polypeptides,
PT useful for diagnosing autosomal recessive spastic ataxia of
PT Charlevoix-Saguenay disease by detecting two point mutations in spastin
PT gene sequence -
XX XX
PS Claim 1; Page -: 76pp; English.
XX XX
CC The present invention describes human and mouse spastin, and mutated
CC human spastin (autosomal recessive spastic ataxia of Charlevoix-Saguenay
CC (ARSACS)) gene sequences (I). The spastin gene has been mapped to
CC chromosome 13q11. (I) have neuroprotective activities and can be used in
CC gene therapy and as a spastin polypeptide agonists. (I), their fragments
CC or their complements can be useful for assaying the presence of a nucleic
CC acid molecule in a sample. (I) is useful for diagnosing or aiding in the
CC diagnosis of an early onset neurodegenerative disease in an individual.
CC The neurodegenerative disease comprises reduced sensory nerve conduction,
CC reduced motor nerve velocity, hypermyelination of retinal nerve fibres,
CC atrophy of upper cerebellar vermis, absence of Purkinje cells and
CC abnormal neuronal lipid storage. (I) can also be used to produce
CC antisense nucleic acids, is useful as molecular weight or chromosome
CC markers, to identify genetic disorders, as hybridisation probes or
CC primers, as an antigen, identify and express recombinant protein for
CC analysis, characterisation or therapeutic use, or as markers for tissues
CC in which the corresponding protein is expressed. Diagnostic methods from
CC the present invention can be used to identify subjects having or at risk
CC of developing a disease or disorder associated with aberrant expression
CC or activity of (I). The assays can be utilised to identify a subject
CC having or at risk of developing a disorder associated with spastin
CC protein or spastin gene expression or activity. The present sequence
CC encodes a mutated human spastin.
CC N.B. The present sequence is not given in the present specification but
CC is derived from the human spastin nucleotide sequence (AAH20174) as
CC stated on page 14.
XX XX
SQ Sequence 12793 BP; 4163 A; 2257 C; 2487 G; 3886 T; 0 other;

```

Query Match 99.28; Score 1390; DB 22; Length 12793;
 Best Local Similarity 99.98; Pred. No. 0;
 Matches 1401; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

```

QY 1 gtagagtaaaactcagagcagctcccaagcgacacaaagccttagaagatgcatcc 60
    |||
DB 5300 gtagagtaaaactcagagcagctcccaagcgacacaaagccttagaagatgcatcc 5359
QY 61 aatgctgttttaacaactctggcagagaattggcgcaagaagaattgacgcagaga 120
    |||
DB 5360 aatgctgttttaacaactctggcagagaattggcgcaagaagaattgacgcagaga 5419

```

```

Oy 121 attaagacatccttaatagcataatcctcttgaaaagaaatgttgaagaagcttctca 180
    |||||
Db 5420 attaagaagcatccttaatagcataatcctcttgaaaagaaatgttgaagaagcttctca 5479
Oy 181 aatgctgataatgcaaaagcgacagaatctgtttgttggatccatacagcatcca 240
    |||||
Db 5480 aatgctgataatgcaaaagcgacagaatctgtttgttggatccatacagcatcca 5539
Oy 241 gttgataagaatatttgaataagtggtgcccattgcaaggccagcactttgtgtac 300
    |||||
Db 5540 gttgataagaatatttgaataagtggtgcccattgcaaggccagcactttgtgtac 5599
Oy 301 aacaacaccacatttacaagaatgatagttagagaattcagaatcttggaaaagcagc 360
    |||||
Db 5600 aacaacaccacatttacaagaatgatagttagagaattcagaatcttggaaaagcagc 5659
Oy 361 aaagaggaaatcctataaacttgacagatagagaatgataatctgtgtatcat 420
    |||||
Db 5660 aaagaggaaatcctataaacttgacagatagagaatgataatctgtgtatcat 5719
Oy 421 atcacagacttgcccatcttatttcttgcaatgacatccctgtgtattttgatccat 480
    |||||
Db 5720 atcacagacttgcccatcttatttcttgcaatgacatccctgtgtattttgatccat 5779
Oy 481 gccagatatgacacaggggcacatccatagtcgccgacgcatgttttagaatttgat 540
    |||||
Db 5780 gccagatatgacacaggggcacatccatagtcgccgacgcatgttttagaatttgat 5839
Oy 541 gcagatttttagacacagatctcagatggttcgagatcttattcttggaacccatttaa 600
    |||||
Db 5840 gcagatttttagacacagatctcagatggttcgagatcttattcttggaacccatttaa 5899
Oy 601 ctggaataattgacaaatggttcgaattccctcttcgtaatgacagaataggaaaagtctg 660
    |||||
Db 5900 ctggaataattgacaaatggttcgaattccctcttcgtaatgacagaataggaaaagtctg 5959
Oy 661 gaatttcgtctgttccagacatcagaagaatggtccagaatcttggacaacactgcgc 720
    |||||
Db 5960 gaatttcgtctgttccagacatcagaagaatggtccagaatcttggacaacactgcgc 6019
Oy 721 tcagatggtggcagaactctcattgttcttaatacacaatggaaaactttatttggaa 780
    |||||
Db 6020 tcagatggtggcagaactctcattgttcttaatacacaatggaaaactttatttggaa 6079
Oy 781 atagataagaatgatactgagcttcaaatgctgattatcagtaaaagggcaaaatccacagat 840
    |||||
Db 6080 atagataagaatgatactgagcttcaaatgctgattatcagtaaaagggcaaaatccacagat 6139
Oy 841 gggagacagattgaaaaggaacaattcagatcctcgttaattgatatggttactaaaag 900
    |||||
Db 6140 gggagacagattgaaaaggaacaattcagatcctcgttaattgatatggttactaaaag 6199
Oy 901 aggcagctcaaaagacataccagttcaacaataactatactatgtaacttggagactc 960
    |||||
Db 6200 aggcagctcaaaagacataccagttcaacaataactatactatgtaacttggagactc 6259
Oy 961 gaaggagaatccttactacgtggtcctaatttgytaataagatacagcttttcaagtatgggaaa 1020
    |||||
Db 6260 gaaggagaatccttactacgtggtcctaatttgytaataagatacagcttttcaagtatgggaaa 6319
Oy 1021 gtactcaaaagtgataatcagctcacaagaacagaatattactctttcccaagcgtgt 1080
    |||||
Db 6320 gtactcaaaagtgataatcagctcacaagaacagaatattactctttcccaagcgtgt 6379
Oy 1081 ggaagtgcgtcgtcattactacaactataaaaaaaccacataggccttcgtttttg 1140
    |||||
Db 6380 ggaagtgcgtcgtcattactacaactataaaaaaaccacataggccttcgtttttg 6439
Oy 1141 cctcttcttttggaagctggtgcacatttcaatgtaaatggtccacttgcacgtgatca 1200
    |||||
Db 6440 cctcttcttttggaagctggtgcacatttcaatgtaaatggtccacttgcacgtgatca 6499

```

```

Oy 1201 gccagaagacactgtgctgcatgataatgagtggtgttccgaagtgcagtgaataac 1260
    |||||
Db 6500 gccagaagacactgtgctgcatgataatgagtggtgttccgaagtgcagtgaataac 6559
Oy 1261 agtttaatgacagcatataatagctcctgcataatg-igaattgctaatacagttaaaaaa 1319
    |||||
Db 6560 agtttaatgacagcatataatagctcctgcataatg-igaattgctaatacagttaaaaaa 6619
Oy 1320 cggtaattccctgtgtcttccaaacttgcagtggttacaagaaccccatattcatgt 1379
    |||||
Db 6620 cggtaattccctgtgtcttccaaacttgcagtggttacaagaaccccatattcatgt 6679
Oy 1380 gtaaaagacactttaagaagt 1401
    |||||
Db 6680 gtaaaagacactttaagaagt 6701

```

RESULT 5

AAH20179
ID AAH20179 standard; DNA; 12793 BP.

AC AAH20179;

DT 09-AUG-2001 (first entry)

XX Human mutated spastin nucleotide sequence SEQ ID NO:12.

KW Human; mouse; spastin; ARSACS; chromosome 13q11; identification;
KW autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;
KW neurodegenerative disease; reduced sensory nerve conduction; diagnosis;
KW reduced motor nerve velocity; hypermyelination of retinal nerve fibre;
KW atrophy of upper cerebellar vermis; absence of Purkinje cell;
KW abnormal neuronal lipid storage; genetic disorder; characterisation; ds.

OS Homo sapiens.
OS Synthetic.

PN MO200129266-A2.

XX 26-APR-2001.

PF 20-OCT-2000; 2000MO-US29130.

XX 20-OCT-1999; 99US-0160588.

PR (UYMC-) UNIV MCGILL.

PA (HOP1-) HOPITAL SAINT- JUSTINE.

XX Hudson TJ, Engert J, Richter A;

DR WPI; 2001-308494/32.

XX New isolated polynucleotide, encoding spastin gene, and polypeptides,
PT useful for diagnosing autosomal recessive spastic ataxia of
PT Charlevoix-Saguenay disease by detecting two point mutations in spastin
PT gene sequence -

XX Claim 1; Page -: 76pp; English.

XX The present invention describes human and mouse spastin, and mutated
CC human spastin (autosomal recessive spastic ataxia of Charlevoix-Saguenay
CC (ARSACS)) gene sequences (1). The spastin gene has been mapped to
CC chromosome 13q11. (1) have neuroprotective activities and can be used in
CC gene therapy and as a spastin polypeptide agonists. (1), their fragments
CC or their complements can be useful for assaying the presence of a nucleic
CC acid molecule in a sample. (1) is useful for diagnosing or aiding in the
CC diagnosis of an early onset neurodegenerative disease in an individual.
CC The neurodegenerative disease comprises reduced sensory nerve conduction,
CC reduced motor nerve velocity, hypermyelination of retinal nerve fibres,
CC atrophy of upper cerebellar vermis, absence of Purkinje cells and
CC abnormal neuronal lipid storage. (1) can also be used to produce
CC antisense nucleic acids, is useful as molecular weight or chromosome
CC markers, to identify genetic disorders, as hybridisation probes or


```

RESULT 7
ABL32158/c
ID ABL32158 standard; DNA; 7990 BP.
XX
AC ABL32158;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 131.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytosstatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW gene; ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP07537.
XX
PR 30-JUN-2000; 2000DE-1032529.
XX
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIC-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful
PT for diagnosis and treatment of diseases associated with abnormal
PT cytosine methylation -
XX
PS Claim 1; SEQ ID NO 131; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention.
XX
SQ Sequence 7990 BP; 2223 A; 102 C; 1567 G; 4098 T; 0 other;

Query Match 2.9%; Score 40.6; DB 24; Length 7990;
Best Local Similarity 49.8%; Pred. No. 0.91;
Matches 103; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

```

```

RESULT 8
ABA59780/c
ID ABA59780 standard; DNA; 531 BP.
XX
AC ABA59780;
XX
DT 01-FEB-2002 (first entry)
XX
DE Human foetal liver single exon nucleic acid probe #8085.
XX
KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
OS Homo sapiens.
XX
PN WO200157277-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00669.
XX
PR 04-FEB-2000; 2000US-0180312.
XX
PR 26-MAY-2000; 2000US-0207456.
XX
PR 30-JUN-2000; 2000US-0608408.
XX
PR 03-AUG-2000; 2000US-0632366.
XX
PR 21-SEP-2000; 2000US-0234687.
XX
PR 27-SEP-2000; 2000US-0236359.
XX
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-483447/52.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human fetal liver -
XX
PS Claim 1; SEQ ID NO 8085; 639pp + sequence listing; English.
XX
CC The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human foetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC foetal liver. The present sequence is a single exon nucleic acid
CC probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at filp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 531 BP; 156 A; 108 C; 107 G; 160 T; 0 other;

Query Match 2.8%; Score 39.6; DB 22; Length 531;
Best Local Similarity 53.2%; Pred. No. 0.51;
Matches 84; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

```


ID ABA28269 standard; DNA: 531 BP.
XX
AC ABA28269;
XX
XX
DT 23-JAN-2002 (first entry)
XX
DE Probe #6735 for gene expression analysis in human heart cell sample.
XX
XX Human; gene expression; heart; microarray; vascular system; probe;
KW cardiovascular disease; hypertension; cardiac arrhythmia;
KW congenital heart disease; ss.
XX
OS Homo sapiens.
XX
PN WO200157274-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00666.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488899/53.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
PT hearts -
XX
PS Claim 1; SEQ ID No 6735; 530bp; English.
XX
XX The present invention relates to single exon nucleic acid probes for
CC measuring human gene expression in a sample derived from human heart. The
CC present sequence is one such probe. The probes may be used for
CC predicting, measuring and displaying gene expression in samples derived
CC from the human heart via microarrays. By measuring gene expression, the
CC probes are useful for predicting, diagnosing, grading, staging,
CC monitoring and prognosing diseases of the human heart and vascular system
CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
CC congenital heart disease.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 531 BP; 156 A; 108 C; 107 G; 160 T; 0 other;

Query Match 2.8%; Score 39.6; DB 22; Length 531;
Best Local Similarity 53.2%; Pred. No. 0.51;
Matches 84; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 1241 ttcgagtgacgtggaataacagtttaatacagcatttaagctcctgcatagtgaatt 1300
DB 507 TCCCAAGTACTGGAACATACAGGTGATGCCACAGTCAATTTTTTTTAAATT 448
QY 1301 gctaaatagcttaaaaaaagcgtatttcctcgtgttcgtatccaacattatcagtggtaca 1360
DB 447 TTTTCATAGACAGAGATCTACTATGTTCACAGTCTCTATAAACAATTATGATATAAAA 388
QY 1361 gaacacccctatcatgtgtgtaagagacattaaaga 1398
DB 387 GAAAAAAGTAATCATCTGAAGTTAAGTCTTAATGA 350

RESULT 10

AAK08051/C
ID AAK08051 standard; DNA: 531 BP.
XX
AC AAK08051;
XX
XX
DT 05-NOV-2001 (first entry)
XX
DE Human brain expressed single exon probe SEQ ID NO: 8042.
XX
XX Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200157275-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00667.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-483446/52.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
PT brains -
XX
PS Example 4; SEQ ID NO: 8042; 650bp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is one of the probes of the
CC invention.
XX
SQ Sequence 531 BP; 156 A; 108 C; 107 G; 160 T; 0 other;

Query Match 2.8%; Score 39.6; DB 22; Length 531;
Best Local Similarity 53.2%; Pred. No. 0.51;
Matches 84; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 1241 ttcgagtgacgtggaataacagtttaatacagcatttaagctcctgcatagtgaatt 1300
DB 507 TCCCAAGTACTGGAACATACAGGTGATGCCACAGTCAATTTTTTTTAAATT 448
QY 1301 gctaaatagcttaaaaaaagcgtatttcctcgtgttcgtatccaacattatcagtggtaca 1360
DB 447 TTTTCATAGACAGAGATCTACTATGTTCACAGTCTCTATAAACAATTATGATATAAAA 388
QY 1361 gaacacccctatcatgtgtgtaagagacattaaaga 1398
DB 387 GAAAAAAGTAATCATCTGAAGTTAAGTCTTAATGA 350

RESULT 11
AAK33925/C
ID AAK33925 standard; DNA: 531 BP.
XX
XX AAK33925;


```

OS Homo sapiens.
XX
PN WO200157272-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00663.
XX
PR 04-FEB-2000; 2000US-0180312.
XX
PR 26-MAY-2000; 2000US-0207456.
XX
PR 30-JUN-2000; 2000US-0608408.
XX
PR 03-AUG-2000; 2000US-0632366.
XX
PR 21-SEP-2000; 2000US-0234687.
XX
PR 27-SEP-2000; 2000US-0236359.
XX
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI: 2001-488897/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human placenta -
XX
PS Claim 25; SEQ ID NO 8336; 654pp; English.
XX
CC The present invention relates to single exon nucleic acid probes (SENP).
XX CC The present sequence is one such probe. The probes are useful for
XX producing a microarray for predicting, measuring and displaying gene
XX expression in samples derived from human placenta. The probes are useful
XX for antenatal diagnosis of human genetic disorders.
XX
SQ Sequence 531 BP; 156 A; 108 C; 107 G; 160 T; 0 other;

Query Match 2.8%; Score 39.6; DB 22; Length 531;
Best Local Similarity 53.2%; Pred. NO. 0.51;
Matches 84; Conservative 0; Mismatches 74; Indels 0; Gaps 0.

OY 1241 ttcgaagtgactggaataacaggttaataacagcatlaataagctccctgcatatgtaatt 1300
DB 507 tcccaagtagtcgggaactgacagtgctgacgtccacacatgcccaagtccaatttttttaatt 448
OY 1301 gctaatcaggttaaaaaacggtatttccctggttcgtgacccaacatcagtggttaca 1360
DB 447 ttttcattagacagacagtgctcactatgtttccacgctcctaaataacattatgtgatpaaaaa 388
OY 1361 gaacacccctattcatgttgtgtaagaagacatttaaga 1398
DB 387 gaaaaaaactaaatcatcctgaagttaagcttttaattga 350

RESULT 14
ID AAK87559/C
XX AAK87559 standard; DNA; 10119 BP.
XX
XX AAK87559;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human Immune/haematopoietic antigen genomic sequence SEQ ID NO.42371.
XX
XX Human; Immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX
XX cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
XX WO200157182-A2.
XX
XX
XX 09-AUG-2001.
XX
XX
XX 17-JAN-2001; 2001WO-US01354.
XX
XX

```

PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 01-NOV-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251858.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 XX
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruden SM;
 XX
 DR WPI: 2001-483426/52.
 XX
 PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
 useful for preventing, diagnosing and/or treating cancers and

PT metastasis -
 XX
 PS Disclosure: SEQ ID NO 42371; 3071bp + Sequence Listing; English.
 XX
 CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
 CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
 CC activity, and can be used in gene therapy and vaccine production. (I)
 CC proteins and polynucleotides may be used in the prevention, diagnosis and
 CC treatment of diseases associated with inappropriate (I) expression. For
 CC example, they may be used to treat disorders associated with decreased
 CC expression by rectifying mutations or deletions in a patient's genome
 CC that affect the activity of (I) by expressing inactive proteins or to
 CC supplement the patient's own production of (I). Additionally, (I)
 CC polynucleotides may be used to produce the secreted (I), by inserting
 CC the nucleic acids into a host cell and culturing the cell to express the
 CC protein. (I) proteins and polynucleotides may be used to prevent,
 CC diagnose and treat immune/haematopoietic-related diseases, especially
 CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
 CC to AAK87694 represent human immune/haematopoietic antigen genomic
 CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
 CC represent sequences used in the exemplification of the present invention.
 XX
 SQ Sequence 10119 BP; 2866 A; 2254 C; 2160 G; 2839 T; 0 other;
 XX
 Query Match 2.88; Score 39.6; DB 22; Length 10119;
 Best Local Similarity 53.28; Pred. No. 1.9;
 Matches 84; Conservative 0; Mismatches 74; Indels 0; Gaps 0;
 QY 1241 ttcaagtgactggaataacaggttaatgacagcattatagcttcgcatatgtaatt 1300
 Db 1842 TCCCAAGTAGCTGACACACAGGTCGATGCCATGCCCATCTTTTATTTTATTT 1783
 QY 1301 gctaatcaggttaaaaaacggtatccctggtctgcatccaacattatcaggttaca 1360
 Db 1782 TTTTCATAGACAGACAGGTCTCACTATGTTTCCAGTCTTAATAACATTATGATTAATAA 1723
 QY 1361 gaacacccctatcatgtgtgtaagaagacatttaaga 1398
 Db 1722 GAAAAAGTAATCATCTGGAAGTTAAGCTTTAATGA 1685
 XX
 RESULT 15
 ID AAK73076 standard; DNA; 12247 BP.
 XX
 AC AAK73076;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:27888.
 XX
 DE Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
 KW cytostatic; gene therapy; vaccine; metastasis; ds.
 XX
 OS Homo sapiens.
 XX
 PN MO200157182-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 17-JAN-2001; 2001WO-US01354.
 XX
 XX 31-JAN-2000; 2000US-0179065.
 XX 04-FEB-2000; 2000US-0180628.
 XX 24-FEB-2000; 2000US-0184664.
 XX 02-MAR-2000; 2000US-0186350.
 XX 16-MAR-2000; 2000US-0189874.
 XX 17-MAR-2000; 2000US-0190076.
 XX 18-APR-2000; 2000US-0198123.
 XX 19-MAY-2000; 2000US-0205515.
 XX 07-JUN-2000; 2000US-0209467.
 XX 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000: 2000US-0215135.
PR 07-JUL-2000: 2000US-0216647.
PR 07-JUL-2000: 2000US-0216680.
PR 11-JUL-2000: 2000US-0217487.
PR 11-JUL-2000: 2000US-0217496.
PR 14-JUL-2000: 2000US-0218290.
PR 26-JUL-2000: 2000US-0220963.
PR 26-JUL-2000: 2000US-0220964.
PR 14-AUG-2000: 2000US-0224518.
PR 14-AUG-2000: 2000US-0224519.
PR 14-AUG-2000: 2000US-0225213.
PR 14-AUG-2000: 2000US-0225214.
PR 14-AUG-2000: 2000US-0225266.
PR 14-AUG-2000: 2000US-0225267.
PR 14-AUG-2000: 2000US-0225268.
PR 14-AUG-2000: 2000US-0225270.
PR 14-AUG-2000: 2000US-0225447.
PR 14-AUG-2000: 2000US-0225757.
PR 14-AUG-2000: 2000US-0225758.
PR 14-AUG-2000: 2000US-0225759.
PR 18-AUG-2000: 2000US-0226279.
PR 22-AUG-2000: 2000US-0226681.
PR 22-AUG-2000: 2000US-0226688.
PR 22-AUG-2000: 2000US-0227182.
PR 23-AUG-2000: 2000US-0227009.
PR 30-AUG-2000: 2000US-0228924.
PR 01-SEP-2000: 2000US-0229287.
PR 01-SEP-2000: 2000US-0229343.
PR 01-SEP-2000: 2000US-0229344.
PR 01-SEP-2000: 2000US-0229345.
PR 05-SEP-2000: 2000US-0229509.
PR 05-SEP-2000: 2000US-0229513.
PR 06-SEP-2000: 2000US-0230437.
PR 08-SEP-2000: 2000US-0230438.
PR 08-SEP-2000: 2000US-0231242.
PR 08-SEP-2000: 2000US-0231243.
PR 08-SEP-2000: 2000US-0231244.
PR 08-SEP-2000: 2000US-0231413.
PR 08-SEP-2000: 2000US-0231414.
PR 08-SEP-2000: 2000US-0232080.
PR 08-SEP-2000: 2000US-0232081.
PR 12-SEP-2000: 2000US-0231968.
PR 14-SEP-2000: 2000US-0232397.
PR 14-SEP-2000: 2000US-0232398.
PR 14-SEP-2000: 2000US-0232399.
PR 14-SEP-2000: 2000US-0232400.
PR 14-SEP-2000: 2000US-0232401.
PR 14-SEP-2000: 2000US-0233063.
PR 14-SEP-2000: 2000US-0233064.
PR 21-SEP-2000: 2000US-0234223.
PR 21-SEP-2000: 2000US-0234223.
PR 21-SEP-2000: 2000US-0234274.
PR 25-SEP-2000: 2000US-0234997.
PR 25-SEP-2000: 2000US-0234998.
PR 26-SEP-2000: 2000US-0235484.
PR 27-SEP-2000: 2000US-0235834.
PR 27-SEP-2000: 2000US-0235836.
PR 29-SEP-2000: 2000US-0236327.
PR 29-SEP-2000: 2000US-0236327.
PR 29-SEP-2000: 2000US-0236367.
PR 29-SEP-2000: 2000US-0236368.
PR 29-SEP-2000: 2000US-0236369.
PR 29-SEP-2000: 2000US-0236370.
PR 02-OCT-2000: 2000US-0236802.
PR 02-OCT-2000: 2000US-0237037.
PR 02-OCT-2000: 2000US-0237037.
PR 02-OCT-2000: 2000US-0237038.
PR 02-OCT-2000: 2000US-0237039.
PR 02-OCT-2000: 2000US-0237040.
PR 13-OCT-2000: 2000US-0239935.
PR 13-OCT-2000: 2000US-0239937.
PR 20-OCT-2000: 2000US-0240960.
PR 20-OCT-2000: 2000US-0241221.
PR 20-OCT-2000: 2000US-0241785.
PR 20-OCT-2000: 2000US-0241786.

PR 20-OCT-2000: 2000US-0241787.
PR 20-OCT-2000: 2000US-0241808.
PR 20-OCT-2000: 2000US-0241809.
PR 20-OCT-2000: 2000US-0241826.
PR 01-NOV-2000: 2000US-0244617.
PR 08-NOV-2000: 2000US-0246474.
PR 08-NOV-2000: 2000US-0246475.
PR 08-NOV-2000: 2000US-0246476.
PR 08-NOV-2000: 2000US-0246477.
PR 08-NOV-2000: 2000US-0246478.
PR 08-NOV-2000: 2000US-0246523.
PR 08-NOV-2000: 2000US-0246524.
PR 08-NOV-2000: 2000US-0246525.
PR 08-NOV-2000: 2000US-0246526.
PR 08-NOV-2000: 2000US-0246527.
PR 08-NOV-2000: 2000US-0246528.
PR 08-NOV-2000: 2000US-0246532.
PR 08-NOV-2000: 2000US-0246532.
PR 08-NOV-2000: 2000US-0246609.
PR 08-NOV-2000: 2000US-0246610.
PR 08-NOV-2000: 2000US-0246611.
PR 08-NOV-2000: 2000US-0246613.
PR 17-NOV-2000: 2000US-0249207.
PR 17-NOV-2000: 2000US-0249208.
PR 17-NOV-2000: 2000US-0249208.
PR 17-NOV-2000: 2000US-0249209.
PR 17-NOV-2000: 2000US-0249210.
PR 17-NOV-2000: 2000US-0249211.
PR 17-NOV-2000: 2000US-0249212.
PR 17-NOV-2000: 2000US-0249213.
PR 17-NOV-2000: 2000US-0249214.
PR 17-NOV-2000: 2000US-0249215.
PR 17-NOV-2000: 2000US-0249216.
PR 17-NOV-2000: 2000US-0249217.
PR 17-NOV-2000: 2000US-0249218.
PR 17-NOV-2000: 2000US-0249244.
PR 17-NOV-2000: 2000US-0249245.
PR 17-NOV-2000: 2000US-0249264.
PR 17-NOV-2000: 2000US-0249265.
PR 17-NOV-2000: 2000US-0249297.
PR 17-NOV-2000: 2000US-0249299.
PR 17-NOV-2000: 2000US-0249300.
PR 01-DEC-2000: 2000US-0250160.
PR 01-DEC-2000: 2000US-0250391.
PR 05-DEC-2000: 2000US-0251030.
PR 05-DEC-2000: 2000US-0251988.
PR 05-DEC-2000: 2000US-0256719.
PR 06-DEC-2000: 2000US-0251479.
PR 08-DEC-2000: 2000US-0251856.
PR 08-DEC-2000: 2000US-0251856.
PR 08-DEC-2000: 2000US-0251869.
PR 08-DEC-2000: 2000US-0251869.
PR 08-DEC-2000: 2000US-0251989.
PR 08-DEC-2000: 2000US-0251990.
PR 11-DEC-2000: 2000US-0254097.
PR 05-JAN-2001: 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.
Rosen CA, Barash SC, Ruben SM:
WPI, 2001-4B3426/52.

Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
useful for preventing, diagnosing and/or treating cancers and
metastasis -

PS Disclosure: SEQ ID NO 27888: 3071pp + Sequence Listing: English.
XX
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins, and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 22, 2002, 06:32:38 ; Search time 108.82 Seconds
(without alignments)
3162.402 Million cell updates/sec

Title: US-09-693-205-7_COPY_5300_6700

Perfect score: 1401
Sequence: 1 gtagcagcgaactaactagagc.....aaagcacttaagaagt 1401

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/prodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/prodata/1/ina/6A_COMB.seq:*
3: /cgn2_6/prodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/prodata/1/ina/6A_COMB.seq:*
5: /cgn2_6/prodata/1/ina/6A_COMB.seq:*
6: /cgn2_6/prodata/1/ina/6A_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	43.2	3.1	7218	1	US-08-232-463-14
2	35	2.5	87350	3	US-08-781-891-79
3	34.2	2.4	423	1	US-08-470-179-142
4	33.8	2.4	305	3	US-08-480-640A-225
5	33.8	2.4	305	4	US-08-686-968C-225
6	33.8	2.4	305	4	US-08-488-237A-225
7	33.8	2.4	2651	1	US-08-462-949-1
8	33.8	2.4	2651	1	US-08-423-764B-1
9	33.6	2.4	4203	2	US-08-866-757-1
10	33.6	2.4	4203	2	US-08-866-757-1
11	32.8	2.3	1007	3	US-08-924-747-3
12	32.8	2.3	1007	4	US-09-247-373B-3
13	32.8	2.3	1007	4	US-09-296-715-3
14	32.6	2.3	1750	3	US-09-120-365-90
15	32.6	2.3	1750	4	US-09-515-039-90
16	32.2	2.3	8920	2	US-08-446-855A-1
17	32.2	2.3	8920	2	US-08-446-855A-1
18	32.2	2.3	8920	4	US-09-150-741-1
19	32.2	2.3	3180	3	US-08-480-662-1
20	32.2	2.3	3180	3	US-08-480-662-1
21	32.2	2.3	3180	3	US-08-918-190-1
22	32.2	2.3	3180	5	US-09-234-232-1
23	32.2	2.3	3180	5	PCT-US96-09927-1
24	32.2	2.3	246240	2	US-08-724-394A-20
25	32.2	2.3	246240	2	US-08-724-394A-21
26	31.8	2.3	2304	1	US-08-724-394A-22
27	31.8	2.3	2304	1	US-08-020-245A-1

28	31.8	2.3	3499	2	US-08-658-665-52	Sequence 52, Appl
29	31.8	2.3	3499	4	US-09-085-273-52	Sequence 52, Appl
30	31.8	2.3	7859	1	US-07-854-596B-4	Sequence 4, Appl
31	31.8	2.3	7859	2	US-08-450-905B-15	Sequence 15, Appl
32	31.8	2.3	7859	3	US-07-982-759F-15	Sequence 15, Appl
33	31.4	2.2	2549	1	US-08-470-720-2	Sequence 2, Appl
34	31.4	2.2	4964	1	US-08-470-720-5	Sequence 5, Appl
35	31.2	2.2	1361	1	US-08-118-469A-4	Sequence 4, Appl
36	31.2	2.2	1361	1	US-08-909-119-4	Sequence 4, Appl
37	31.2	2.2	1400	4	US-09-117-257-10	Sequence 10, Appl
38	31.2	2.2	1400	4	US-08-945-476-10	Sequence 10, Appl
39	31.2	2.2	1400	4	US-09-489-352-10	Sequence 10, Appl
40	31.2	2.2	1498	1	US-07-965-668A-1	Sequence 1, Appl
41	31.2	2.2	1498	2	US-08-950-433-1	Sequence 1, Appl
42	31.2	2.2	1498	3	US-09-186-287-1	Sequence 1, Appl
43	31.2	2.2	2653	2	US-08-589-711-1	Sequence 1, Appl
44	31.2	2.2	2653	4	US-09-221-938-1	Sequence 1, Appl
45	31.2	2.2	2653	4	US-08-945-476-7	Sequence 7, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14/C
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHERFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300,6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: PTZ9PL-Fls
US-08-232-463-14

Query Match	Similarity	3.1%	Score 43.2	DB 1	Length 7218
Best Local	5.8%	Pred. 0.016			
Matches 21	Conservative 188	Mismatches 151	Indels 0	Gaps 0	
QY	43	ttagaagaatagcatgcacatgctcgttttacaacactgycagcaaatltggcga	102		
Db	1479	TTACTATCTATGCAAGTAAAGACATACAGAAATTGGTACRRRRRRRRRR	1420		
QY	103	gaaaaattgacccagcagaattaaagcatccttaatgcatatcctctgaaga	162		
Db	1419	RR	1360		
QY	163	ttgaagagctctccaatgtcgtatgtagtaagggcagcagaatcgtttgtt	222		
Db	1359	RR	1300		
QY	223	gacctagacagcatccagttgatgaatattgtatgaagtgggcccatgtc	282		
Db	1299	RR	1240		
QY	283	ccagcatcttggtgtacacaacacagccattacagaagatgtatgagaat	342		
Db	1239	RR	1180		
QY	343	aatctggaaaaagcagcaagaaggaatccttataaacctgcagcatgtga	402		
Db	1179	RR	1120		

US-08-781-891-79/c
 Sequence 79, Application US/08781891
 Patent No. 6090620
 GENERAL INFORMATION:
 APPLICANT: Fu, Ying-Hui
 APPLICANT: Yu, Chang-En
 APPLICANT: Oshima, Junko
 APPLICANT: Mulligan, John T.
 APPLICANT: Schellenberg, Gerald D.
 TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
 NUMBER OF SEQUENCES: 209
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: SEED and BERRY LLP
 STREET: 6300 Columbia Center, 701 Fifth Avenue
 CITY: Seattle
 STATE: Washington
 COUNTRY: USA
 ZIP: 98104-7092
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/781,891
 FILING DATE: 27-DEC-1996
 CLASSIFICATION: 800
 ATTORNEY/AGENT INFORMATION:
 NAME: Mr. 6090620tenburg Ph.D., Carol
 REGISTRATION NUMBER: 39,317
 REFERENCE/DOCKET NUMBER: 240052.419
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (206) 622-4900
 TELEFAX: (206) 682-6031
 INFORMATION FOR SEQ ID NO: 79:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 87350 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 US-08-781-891-79

Query Match	Similarity	2.5%	Score 35	DB 3	Length 87350
Best Local	Similarity	51.6%	Pred. No. 8.4		
Matches	80	Conservative	0	Mismatches	75
				Indels	0
				Gaps	0
Oy	819	aglaaaggccaatacacagatcgtagacagatltgaagaacatltcalgcactcgt	878		
Db	49881	ACTGAGGGGGAAAAACCGAAGAAGATCTCGAATATAAAGAAAAAATTATACAGCTTC	49822		
Oy	879	aatgcatactgcttacttaaaaaggcagctcaagaacataccagttccaacaataccta	938		
Db	49821	AGTTAATAGTGTGCTCCTTGTTGGTAAAAATAAATAATGTATATATAAATAAAAAAGAAAT	49762		
Oy	939	tactatgatacttgagagactctgaaggaacactta	973		
Db	49761	TTTTTAATCTAAAAAACGTTTTAAATATATTTTAA	49727		

```

1      RESULT      3
2      US-08-470-179-142
3      : Sequence 142, Application US/08470179
4      : Patent No. 5645994
5      :
6      : GENERAL INFORMATION:
7      : APPLICANT: Huang Ph.D, Wai Mun
8      : TITLE OF INVENTION: Method and Compositions for
9      : TITLE OF INVENTION: Identification of Species in a Sample
10     : NUMBER OF SEQUENCES: 207
11     : CORRESPONDENCE ADDRESS:
12     : ADDRESSEE: Task, Britl and Rossa
13     : STREET: P.O. Box 2550
14     : CITY: Salt Lake City
15     : STATE: Utah
16     : COUNTRY: USA
17     : ZIP: 84110
18     :
19     : COMPUTER READABLE FORM:
20     : MEDIUM TYPE: Floppy disk
21     : COMPUTER: IBM PC compatible
22     : OPERATING SYSTEM: PC-DOS/MS-DOS
23     : SOFTWARE: Patent In Release #1.0, Version #1.30
24     : CURRENT APPLICATION DATA:
25     : APPLICATION NUMBER: US/08/470,179
26     : FILING DATE:
27     : CLASSIFICATION: 435
28     : ATTORNEY/AGENT INFORMATION:
29     : NAME: Sweigert Ph.D, Susan E.
30     : REGISTRATION NUMBER: 36,289
31     : REFERENCE/DOCKET NUMBER: 2601
32     : TELECOMMUNICATION INFORMATION:
33     : TELEPHONE: 801-532-1922
34     : TELEFAX: 801-531-9168
35     : INFORMATION FOR SEQ ID NO: 142:
36     : SEQUENCE CHARACTERISTICS:
37     : LENGTH: 423 base pairs
38     : TYPE: nucleic acid
39     : STRANDEDNESS: double
40     : TOPOLOGY: not relevant
41     : MOLECULE TYPE: DNA (genomic)
42     : HYPOTHETICAL: NO
43     : ANTI-SENSE: NO
44     : ORIGINAL SOURCE:
45     : ORGANISM: Mycoplasma arthritidis
46     :
47     : US-08-470-179-142

```

[illegible]

LENGTH: 305 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: N
ANTI-SENSE: N
US-08-488-237A-225

Query Match
Best Local Similarity 2.4%; Score 33.8; DB 4; Length 305;
Matches 95; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

QY 844 gacagatgaaaggaacatttcacatcgtatgatgattgattactaaagag 903
DB 280 GATCATAGAAACAGACATATGATGTTCTTATCGAATATAACATTTCCATGAGATA 221
QY 904 cagctcaagacataccagttcaacaataacctatactatgatactgagactgaa 963
DB 220 AATAACCAAAATCCAAACATTAATATCTGTGCTGATTTCTTATTTATCTCTAGT 161
QY 964 ggaatctactacgtggtcatttgtaagtcagcgtttcaagatgagaaagta 1023
DB 160 ACATACGATGATGCTTATTATAAAACACATCAACTATCATGAAATAGGTATATTA 101
QY 1024 tctaaagtcatc 1040
DB 100 TTTAAATCTCTATTC 84

RESULT 7
US-08-949-1
Sequence 1, Application US/08462949
Patent No. 5606022

GENERAL INFORMATION:
APPLICANT: Rasmussen, Beth Ann
TITLE OF INVENTION: Cloning and Identification of a Two
TITLE OF INVENTION: Component Signal Transducing Regulatory System from
TITLE OF INVENTION: Bacteroides Fragilis
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dady & Dady P.C.
STREET: 805 Third Avenue
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10022

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/462,949
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/023,764
FILING DATE:

ATTORNEY/AGENT INFORMATION:
NAME: Robinson, Joseph R.
REGISTRATION NUMBER: 33,448
REFERENCE/DOCKET NUMBER: 0646/1B024-US1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-527-7700
TELEFAX: 201-753-6237
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2651 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-462-949-1

Query Match
Best Local Similarity 2.4%; Score 33.8; DB 1; Length 2651;
Matches 71; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 746 tcttaatcacatgaaataatttcatttgyaataagatagatcagagctctaa 805
DB 261 TTCTATGAAAAGTCACATCTGATATATGATATGAGGCTTCTTCCTTCTGA 320
QY 806 atgtgctgtaactgaagaagcaaatcacagatgagacagattgaaagaaacaat 865
DB 321 GTTGCGCTTTTTCAGATGAGCTACATCGAAGAAATGCGAGATGCCATGAACAAAT 380
QY 866 tcatgcatctgt 878
DB 381 TTAATACATCCGT 393

RESULT 8
US-08-023-764B-1
Sequence 1, Application US/08023764B
Patent No. 5679540

GENERAL INFORMATION:
APPLICANT: Rasmussen, Beth Ann
TITLE OF INVENTION: Cloning and Identification of a Two
TITLE OF INVENTION: Component Signal Transducing Regulatory System from
TITLE OF INVENTION: Bacteroides Fragilis
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: American Cyanamid Company
STREET: One Cyanamid Plaza
CITY: Wayne
STATE: New Jersey
COUNTRY: United States
ZIP: 07470-8426

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/023,764B
FILING DATE: 26-FEB-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Barnhard, Elizabeth M.
REGISTRATION NUMBER: 31,088
REFERENCE/DOCKET NUMBER: 31,658-00
TELECOMMUNICATION INFORMATION:
TELEPHONE: (201)831-3246
TELEFAX: (201)831-3305
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2651 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-023-764B-1

Query Match
Best Local Similarity 2.4%; Score 33.8; DB 1; Length 2651;
Matches 71; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 746 tcttaatcacatgaaataatttcatttgyaataagatagatcagagctctaa 805

Db 261 TTGCTATGAAAGTCAACATCTGATATAGCATATATATGAGCTTCTTCCTTTCGA 320
Qy 806 atggtccgtatttcagtaagggcaaatcacagatggaagacattgaaaggaacaat 865
Db 321 GTTGCGCTATTATTCAGAGTGAGTACATCGAAGAAATGGTGAGATGGGTAAAGAACAT 380
Qy 866 ttcatgcatctgt 878
Db 381 TTAATACATCCGT 393

RESULT 9

US-08-866-757-1/c
; Sequence 1, Application US/08866757
; Patent No. 5838716
; GENERAL INFORMATION:
; APPLICANT: ELSHOURBAGY, NABIL A
; APPLICANT: LI, XIAOTONG
; APPLICANT: BERGSM, DERK J
; TITLE OF INVENTION: NOVEL 7TM RECEPTOR (H2CAA71)
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: RATNER & PRESTIA
; STREET: P.O. BOX 980
; CITY: VALLEY FORGE
; STATE: PA
; COUNTRY: USA
; ZIP: 19482
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/866,757
; FILING DATE: 30-MAY-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: PRESTIA, PAUL F
; REGISTRATION NUMBER: 23,031
; REFERENCE/DOCKET NUMBER: GH-70055
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-407-0700
; TELEFAX: 610-407-0701
; TELEX: 846169
; INFORMATION FOR SRO ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4203 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-866-757-1

Query Match 2.4%; Score 33.6; DB 2; Length 4203;
Best Local Similarity 51.3%; Pred. No. 6.3;
Matches 78; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

Qy 524 tgttagagattcgatgcagatttaggacacagttctcagatgctcgtatctatc 583
Db 2867 TGTCCTGAGATTAAAGATGAACATTTAGGTAAATTAGCAGATATTTCTTCAGTTATA 2808
Qy 584 tgggaaccatttaactgataatgacaaatgctcagattcctccttcgtaatgacg 643
Db 2807 GAAATACACATTTTAAATATATTTTAAATAATTAACAGCTGTCTTATCTTGTCG 2748
Qy 644 aaatgcaaaagtctcgaaattcgtctgt 675

Db 2747 TTAATACAAATAGTATTAATTTGCTTCTT 2716

RESULT 10

US-09-153-593-1/c
; Sequence 1, Application US/09153593A
; Patent No. 6174994
; GENERAL INFORMATION:
; APPLICANT: ELSHOURBAGY, NABIL A
; APPLICANT: LI, XIAOTONG
; APPLICANT: BERGSM, DERK J
; TITLE OF INVENTION: NOVEL 7TM RECEPTOR (H2CAA71)
; FILE REFERENCE: GH-70055-1
; CURRENT APPLICATION NUMBER: US/09/153,593A
; CURRENT FILING DATE: 1998-09-15
; EARLIER APPLICATION NUMBER: 08/866,757
; EARLIER FILING DATE: 1997-05-30
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 4203
; TYPE: DNA
; ORGANISM: HOMO SAPIENS
; US-09-153-593-1

Query Match 2.4%; Score 33.6; DB 4; Length 4203;
Best Local Similarity 51.3%; Pred. No. 6.3;
Matches 78; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

Qy 524 tgttagagattcgatgcagatttaggacacagttctcagatgctcgtatctatc 583
Db 2867 TGTCCTGAGATTAAAGATGAACATTTAGGTAAATTAGCAGATATTTCTTCAGTTATA 2808
Qy 584 tgggaaccatttaactgataatgacaaatgctcagattcctccttcgtaatgacg 643
Db 2807 GAAATACACATTTTAAATATATTTTAAATAATTAACAGCTGTCTTATCTTGTCG 2748
Qy 644 aaatgcaaaagtctcgaaattcgtctgt 675
Db 2747 TTAATACAAATAGTATTAATTTGCTTCTT 2716

RESULT 11
US-08-924-747-3
; Sequence 3, Application US/08924747
; Patent No. 6063570
; GENERAL INFORMATION:
; APPLICANT: MCGONIGLE, BRIAN
; APPLICANT: O'KEEFE, DANIEL
; TITLE OF INVENTION: SOYBEAN GLUTATHIONE-S-TRANSFERASE
; TITLE OF INVENTION: ENZYMES
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
; STREET: 1007 MARKET STREET
; CITY: WILMINGTON
; STATE: DELAWARE
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.50 INCH
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: MICROSOFT WORD FOR WINDOWS 95
; SOFTWARE: MICROSOFT WORD VERSION 7.0A
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/924,747
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: FLOYD, LINDA AXAMETHY
; REGISTRATION NUMBER: 33,692
; REFERENCE/DOCKET NUMBER: CL-1108

TELECOMMUNICATION INFORMATION:
TELEPHONE: 302-892-8112
TELEFAX: 302-773-0164
INFORMATION FOR SEQ. ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1007 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE: SOYBEAN
TISSUE TYPE: IMMEDIATE SOURCE:
CLONE: SSM.PK0026.G11
US-08-924-747-3

Query Match 2.3%; Score 32.8; DB 3; Length 1007;
Best Local Similarity 58.0%; Pred. No. 6;
Matches 58; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

OY 930 aataacctatactatgatactgagactctgaagaaatcttactacgtgctaattg 989
DB 380 ACTGAACACTACATGGGAGAAAGTTGGCCCTGATGAAAACTTCCTGGGCCCAAGTAT 439

OY 990 taatagatcaggctttcaagtatgagaagaatctctaa 1029
DB 440 AATTAGAAAGGCTTTAAAGCACTGAAAGCTATTGAAA 479

RESULT 12
US-09-247-373B-3
Sequence 3, Application US/09247373B
Patent No. 6168954
GENERAL INFORMATION:
APPLICANT: MCGONIGLE, BRIAN
APPLICANT: O'KEEFE, DANIEL
TITLE OF INVENTION: SOYBEAN GLUTATHIONE-S-TRANSFERASE ENZYMES
FILE REFERENCE: CL-1108-A
CURRENT APPLICATION NUMBER: US/09/247,373B
CURRENT FILING DATE: 1999-02-10
PRIOR APPLICATION NUMBER: 08/924,747
PRIOR FILING DATE: 1997-09-05
NUMBER OF SEQ. ID NOS: 56
SOFTWARE: Microsoft Office 97
SEQ. ID NO 3
LENGTH: 1007
TYPE: DNA
ORGANISM: SOYBEAN
US-09-247-373B-3

Query Match 2.3%; Score 32.8; DB 4; Length 1007;
Best Local Similarity 58.0%; Pred. No. 6;
Matches 58; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

OY 930 aataacctatactatgatactgagactctgaagaaatcttactacgtgctaattg 989
DB 380 actgaactaatattgggagaagtgtgcccctgataaacttccttggtgccaagtat 439
OY 990 taatagatcaggctttcaagtatgagaagaatctctaa 1029
DB 440 aattagaagaggctttaaagcactggaagaagctattgaaa 479

RESULT 13
US-09-296-715-3
Sequence 3, Application US/09296715
Patent No. 6171839
GENERAL INFORMATION:
APPLICANT: MCGONIGLE, BRIAN

APPLICANT: O'KEEFE, DANIEL
TITLE OF INVENTION: SOYBEAN GLUTATHIONE-S-TRANSFERASE
TITLE OF INVENTION: ENZYMES
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: E.I. DU PONT DE NEMOURS AND COMPANY
STREET: 1007 MARKET STREET
CITY: WILMINGTON
STATE: DELAWARE
COUNTRY: UNITED STATES OF AMERICA
ZIP: 19898
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.50 INCH
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: MICROSOFT WORD FOR WINDOWS 95
SOFTWARE: MICROSOFT WORD VERSION 7.0A
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/296,715
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: FLOYD, LINDA AXAMETHY
REGISTRATION NUMBER: 33,692
REFERENCE/DOCKET NUMBER: CL-1108
TELECOMMUNICATION INFORMATION:
TELEPHONE: 302-892-8112
TELEFAX: 302-773-0164
INFORMATION FOR SEQ. ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1007 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE: SOYBEAN
TISSUE TYPE: IMMEDIATE SOURCE:
CLONE: SSM.PK0026.G11
US-09-296-715-3

Query Match 2.3%; Score 32.8; DB 4; Length 1007;
Best Local Similarity 58.0%; Pred. No. 6;
Matches 58; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

OY 930 aataacctatactatgatactgagactctgaagaaatcttactacgtgctaattg 989
DB 380 ACTGAACACTACATGGGAGAAAGTTGGCCCTGATGAAAACTTCCTGGGCCCAAGTAT 439
OY 990 taatagatcaggctttcaagtatgagaagaatctctaa 1029
DB 440 AATTAGAAAGGCTTTAAAGCACTGAAAGCTATTGAAA 479

RESULT 14
US-09-120-365-90
Sequence 90, Application US/09120365
Patent No. 6103514
GENERAL INFORMATION:
APPLICANT: Natori, Shunji
TITLE OF INVENTION: NEW PROTEASE
FILE REFERENCE: 32290-144749
CURRENT APPLICATION NUMBER: US/09/120,365
CURRENT FILING DATE: 1998-07-22
EARLIER APPLICATION NUMBER: JP 9-333 474
PRIOR FILING DATE: 1997-11-18
NUMBER OF SEQ. ID NOS: 101
SOFTWARE: PatentIn Ver. 2.0
SEQ. ID NO 90
LENGTH: 1750
TYPE: DNA

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 22, 2002, 04:26:55 ; Search time 3530.57 Seconds
(Without alignments)
2963.620 Million cell updates/sec

Title: US-09-693-205-7_COPY_1_500
Perfect score: 500
Sequence: 1 atgattcagaagacatc.....gcagaattgttaatacaaa 500

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenBank: *
1: gb_ba: *
2: gb_htg: *
3: gb_in: *
4: gb_ov: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_sts: *
12: gb_sy: *
13: gb_un: *
14: gb_vl: *
15: em_ba: *
16: em_fun: *
17: em_hum: *
18: em_in: *
19: em_mu: *
20: em_om: *
21: em_or: *
22: em_ov: *
23: em_pat: *
24: em_ph: *
25: em_pl: *
26: em_ro: *
27: em_sts: *
28: em_un: *
29: em_vl: *
30: em_htg_hum: *
31: em_htg_inv: *
32: em_htg_other: *
33: em_htgo_inv: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
------------	-------	-------	-------	--------	----	----	-------------

1	500	100.0	12793	6	AX119931	AX119931 Sequence
2	500	100.0	12793	9	AF193556	AF193556 Homo sapi
3	500	100.0	92693	9	AF157766	AF157766 Homo DNA
4	499	99.8	99819	2	AC079761	AC079761 Homo sapi
5	344	68.8	174140	2	AC069017	AC069017 Mus muscu
6	331.2	66.2	11493	10	AF193557	AF193557 Mus muscu
7	328.6	65.7	11492	6	AX119933	AX119933 Sequence
8	42.2	8.4	12029	3	AE001381	AE001381 Plasmodu
9	42.2	8.4	116127	3	AC096291	AC096291 Rattus no
10	41.6	8.3	110960	3	CER39A1A	ALD31633 Caenorhab
11	41.6	8.3	160624	9	AC060835	AC060835 Homo sapi
12	41.6	8.3	168982	2	AC020669	AC020669 Homo sapi
13	41.6	8.3	206235	2	AC107486	AC107486 Homo sapi
14	41.4	8.3	49623	3	AC087232	AC087232 Caenorhab
15	40.2	8.0	39419	3	CER501C7	AC084475 Caenorhab
16	40.2	8.0	161404	9	AP002376	AP002376 Homo sapi
17	39.8	8.0	157637	2	OSJN00122	ALD06590 Oryza sat
18	39.6	7.9	40242	8	SPAC1F5	268136 S. pombe ch
19	39.6	7.9	95796	2	AC094350	AC094350 Rattus no
20	39.6	7.9	169541	9	AL136178	AL136178 Human DNA
21	39.6	7.9	175485	2	AC073802	AC073802 Mus muscu
22	39.6	7.9	186496	2	AC020573	AC020573 Homo sapi
23	39.4	7.9	98913	8	AF069442	AF069442 Arabidops
24	39.4	7.9	176749	9	AC096748	AC096748 Homo sapi
25	39.4	7.9	199200	2	AC109830	AC109830 Homo sapi
26	39.4	7.9	199615	8	ATCHR1V7	AL161495 Arabidops
27	39.2	7.8	106601	9	AL357352	AL357352 Human DNA
28	39.2	7.8	160750	9	HSJ177110	AL080284 Human DNA
29	39.2	7.8	165660	2	AC027194	AC027194 Homo sapi
30	39.2	7.8	181298	2	AC008333	AC008333 Drosophi1
31	39.2	7.8	184003	2	AC011058	AC011058 Homo sapi
32	39.2	7.8	93443	8	ATF2206	AL050300 Arabidops
33	39.2	7.8	100028	8	ATF3C22	AL353912 Arabidops
34	39.2	7.8	146802	9	AC025186	AC025186 Homo sapi
35	39.2	7.8	178113	2	AC099516	AC099516 Homo sapi
36	39.2	7.8	190044	9	AC021842	AC021842 Homo sapi
37	38.8	7.8	825	3	AF206632	AF206632 Plasmodu
38	38.8	7.8	914	3	PRRESA	X55124 P. falciparu
39	38.8	7.8	21137	8	AC093617	AC093617 Homo sapi
40	38.8	7.8	104386	8	ATT32A17	AL161813 Arabidops
41	38.8	7.8	131949	2	AP003341	AP003341 Oryza sat
42	38.8	7.8	149313	2	AC023548	AC023548 Homo sapi
43	38.8	7.8	158084	8	AP003492	AP003492 Oryza sat
44	38.8	7.8	199362	8	ATCHR1V24	AL161512 Arabidops
45	38.4	7.7	2069	9	BC007641	BC007641 Homo sapi

ALIGNMENTS

RESULT	1	AX119931	12793 bp	DNA	linear	PAT 11-MAY-2001
LOCUS	AX119931	Sequence	1 from Patent W00129266.			
DEFINITION	AX119931					
ACCESSION	AX119931					
VERSION	AX119931.1	GI:14036678				
KEYWORDS						
SOURCE		human.				
ORGANISM		Homo sapiens				
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;						
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.						

REFERENCE
1 (bases 1 to 12793)
Hudson, T.J., Engert, J. and Richter, A.
Identification of arsacs mutations and methods of use therefor
Patent: WO 0129266-A 1 26-APR-2001;
MCGILL UNIVERSITY (CA) ; Hopital Sainte-Justine (CA)
Location/Qualifiers
FEATURES
SOURCE
1..12793
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT	4163 a	2256 c	2487 g	3887 t
ORIGIN				

Query Match 100.0%; Score 500; DB 6; Length 12793;
 Best Local Similarity 100.0%; Pred. No. 6.9e-101;
 Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 atgattcagaagaacacatgctactcagctgcagctcttaaatcagaacgattgcag 60
DB 1 ATGATTACAGGAAGACCATGTACTCAGTCGACGCTTCTAAATCCAGACGATTGCACG 60
QY 61 tctatcagaagaatgataatcatctctgctgcgagagaatgattgattcaatgta 120
DB 61 TCTATACAGGAAGTAATGATACATTTCTGCTGCCGAGGAATGATTGATTCTCAATGTA 120
QY 121 tccattgataaacaagaatacaccatcgttttcaaggcttaagaatgattggaanaa 180
DB 121 TCATTGATGAGAAACAGAAATCACCCATCTGTTTCATGCTTAGATGCTTGGAANA 180
QY 181 tcttatatacatcttccagaagattgacttattatgatagatagcacttaccacag 240
DB 181 TCTTATATACATTTTTCAGAGGATTGACTTATTATTGATGAGATGCACCTATCCCCAG 240
QY 241 aactatactcagaagaagtgtaacatgctggaacatcattagactcaggatccatcgt 300
DB 241 AACTATACTAGAGAGAGTCAGACATGTGTGGAACCTAGACTCAGGATTCATCGATT 300
QY 301 agcattttagaagatgaatctgaagacagcttcagaattttagacagacattgtaca 360
DB 301 AGCATTTTAGAGATGAATCTGACGACAGCTTCCGAATTTTTCAGACATGTTGACA 360
QY 361 aaaaacttgagaaggtctgtccttaaaaaataagatgatacctatacaacatccgttataa 420
DB 361 AAAACTTGAGAGGCTTGTCTTAAAAAATGATGATCATATACACATCCGCTTATTAA 420
QY 421 aaatatattatcaccatccattaccagtgctgttttgcaataatgagaagaatgcatc 480
DB 421 AAATATATTCATTCACCATTACCAAGTGTGTTTGCAGATATGAGAGAGATGCATT 480
QY 481 gcagaaatgtgtaacaa 500
DB 481 GCAGAAATGTCTGATTCAM 500

RESULT 2
AF193556 12793 bp DNA linear PRI 07-FEB-2000
LOCUS Homo sapiens saccin (SACS) gene, complete cds.
DEFINITION AF193556
ACCESSION AF193556
VERSION AF193556.1 GI:6907041
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 12793)
Engel, J.C., Berube, P., Mercier, J., Dore, C., Lepage, P., Ge, B.,
Bouchard, J.P., Mathieu, J., Melancon, S.B., Schalling, M.,
Lander, E.S., Morgan, K., Hudson, T.J. and Richter, A.
ARSAACS, a spastic ataxia common in northeastern Quebec, is caused
by mutations in a new gene encoding an 11.5-kb ORF
Nat. Genet. 24 (2), 120-125 (2000)
2 (bases 1 to 12793)
Engel, J.C., Berube, P., Dore, C., Lepage, P., Ge, B.,
Richter, A.
Direct Submission
Submitted (08-Oct-1999) Genome Centre, Montreal General Hospital,
1650 Cedar Ave., Montreal, QC H3G 1A4, Canada
Location/Qualifiers
1. 12793
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="13"
/map="between D13S232 and D13S292"

```

mRNA
 gene
 cds

```

1. 12793
/gene="SACS"
/product="saccin"
1. 12793
/gene="SACS"
/gene="SACS"
77. 11566
/note="molecular chaperone"
/codon_start=1
/product="saccin"
/db_xref="GI:6907042"
/translacion="MNTFMPGRELIVQWYFPDENRNHPSVSLKMKWNLVHIFSEDL
TLPEDEMLPRTILLEGOTCEVLEIRLIPSLVILDESEOKNLALMLNLIIMLYSMOIPIS
KIDASTOHLPIKKYTHSP.PSAVIOIMEMPKLOKONITSLIPKIDAKLQKGFALST
TDSSEKERRIIOGLAIFKRIINHSDDGISSTYTLKGGKVLHHTAKLPADRLSISVTD
SSDEATIRLANMLKIDOLKTSCLKLVLKIOENAFYSHEEVTLQMLVFNLSLAKNE
NPVLEMLPPLKFIQISOEOMVAGELFDDIDVLDLFCNEBGTYVPPSPDIL
HSIROIGLKNBASLKEKDVQVAKKEALDVGACPDQDLAKKAKTLLVLKNNHTLL
OSSEGKMTLKKIKWVPACKERPKNPSLWKGDLCLCAPMDGVGHALILGSLP
LVSIVHNLKALGIETPKSLSAVIAKIFKIVDVMYSKPSDEDDYQFOHILLEIYGF
MDHLNKGDSFALKRPWYWTCKKRCPLAQAIVKPLHIDLOPYLHNYKTYAKRHO
LFRVGSIEELTDHISMTVQIKLYLSDSDLSDESKQNLALMLNLIIMLYSMOIPIS
PNTVPPIHHSKNPSKLMKPIHBCCYDIDVDLNDLDESVBPIILVHDIEMKTAIE
MLKPCILSTRILIPENMGFOSGOREPLIVIRIKNIIIEEYPSVDIEFELKNDADANA
TECSPLIDMRNMDIRENLDPGMACHGALMSPNNSOFSDSDPVNITLQNSLKG
ENDVGKREGFESVYHTDIPILMSREPMMDPNHTSKHKKNSNGITIKMCK
OQRKLRFPNOKRPFIDVRCQOLPLYEARYSTNGTLFRLSFTODEAKKSEVSTCY
NTADIVSLDEFSLGHRLIFQSVKSMTLKYLETPSLAODIVILIKKSCSK
ALNTVPYSLVLEAKALMKTKCSSNKKLPSDEPSQIIQTVBEFHFHRRIDLOSP
LFRGPDDPALFEMAKSGQSKPSDELSDQKTEVCTMTLCTMDGELKFLSESG
RRLGLVPCGAVGVQSLIDOKMTWPHIEVFCYPLRIKGTGLPVHINCGFVNSR
KETIMDTKGRMNTFMRHVIYKAYLQVLSVLDLATSGLMDTYTAYAWPDDVDIV
DESVTCQGFETDIAHGKGLTVPFSDGSSVSKMKNVRLDSDILKRDVGSAAKTF
LKLYKTGSKNLCAVELPSSVKLFEEBAGKQILLENTSEKOFSESEFPNIOLEA
ELNDPLMIFVNLKRYDEFGVLTVPDIPCSLGHPLVPSRLIHEGRVAKLPDLKD
GRPPYSTODYLNPILIKLVQLMKADLILMDMDLRAVSAVAINKSDHVAACLRSS
ILSLIDEKLRIDPRAKDDPAKYOTIRFLPFLPAGFSLDKMKNSEFKEPTEMAIR
LYAEHODIVLCILOPILNENSHSFRGGGVSILVKEFGLIGLKKPKYDVLVINOLEAVK
SYVDGTLIVQENTNACRYLHIALMONETTKSTIDKLPPSFILVENAYVSEKYS
VHNFENAPILVOLPNKYNNEFELFETVNGROSCVEDALVESIDORGRKQITE
ENQOLCRAPIISEGIMSLIRKKOPECENKNGKILPLPTNLMLPAKSLCNDPWLK
KDTTVYXCHADIDPREYAVKLGAVPKRKHLEARYASVNCFTTLEGEQAKELSRIS
ILNAVSEKEMLEKELLONADADAKATECFEDYDRONVHDIFPDNAKALOGPALCVN
NOPTEDDVGTONIGKGTKEGNPRTYGQIGTFNSYHTHTDCPSPTSGNDICIDP
HARYARGATISISGRMRDADAPRQFOSVLDLITGTHKLDNCMFRPLNAEBA
KVEISSVPASDRMVOQLDKLRSDSGDELMLFNLHNEKLSICEIDSTGALNLYSVK
GKLTIDDLRLKROPHASVIDSVTKRQLKDIPOQLTYYMTDDESBGNTLWILCNIS
GFSSMEKVSYSIAHKNODITLFPFGVNAACITLHNYKKPRAFCFLPLSEGLPEF
VNGHFAIDSAARRLMDRDNGVGRSDMNSLMTALALAPAVELLIOKKRYPGSDPT
LSVIONTPHVVKDLTKKPLFFPVRNIDLOPLCYLVALVNCIHEMDKRLIPVRA
PNIDGSDLSHAVITITMNSTSKTRPFPNLLODELQHKANADYITTRKYAENVY
RLKHLILEGFNLVYNCDETANLIFHCLIDIDIDVSTVPADISLMTSSPPTNCH
GKLPRLQOQNLKLFHSLKLVLYCERKDAENIEVEGLPLTLTSSLOTPKARRK
FLTYTHIELIPSRDLPMNTLYLKYSNILNLCAKKFEJDSFDDLSSVLRKTKYS
CTKWNKNESESMLKNAWHFISESVKDEQCEPFLFIDVVDLTLDMALPRTKFLV
SNAOIVPBGDVLPLISIMHIAVPPNAGOSVPHALMKACIOALANKIKOSKSAVP
LISCHRANITESPISILKALHMYQTSITFAREKIVENDFEXALLATPNCNLHMLSODI
KILKSLPCYKYSISGRVYSIGKFGCTVLTLSISAEVEKATQSSSSAFLEKJLHLEL
YEVIGCVPDVNDLEVYLKHLPIENLSDYDKLHLIYLLKRRLSAAEELSIKQOLDEK
LESILLIHDANSRLKOKAHEYDRTVVFEEVMLPEKFLIPDPFKKLEOLIKRPNHVF
MTSWVEHNLIGLTYLLSOOLLOFAKEISVRANTENWMSKETQNTYNDLHHIIPER
MOLLSGNFLKEKLSLPIELCPERAPAFIRPHPOYOVNGTLPILKNGAGVNRKFOC
DVIQILMTSCPIIPKATPLSTKEQSGSDGDPQOLBOVLNMLNVLNLPPLDVYVINC
RVNCTNTTTLDEENKTRAKVLRISYIELSEKKEFRQLGVAVFVABDWKLLKLEBE
VYNLEIESDFKPYLYKLPLELTFHQLFHLKLEDEIISIKQVVEVLSRIFKNSBKQ
LDPENMRKYAVRVSGLFRLSNDSVKRSVDSLEVRDLATLVPQDRLVSSLSIVDE
LAPHYKRIQONIGVOMVLDOSCYLGKDHCFHKLILVLPQKLRPLRLSILLEODE
EMPKVOCFALCSIOGRLOLILSSQFPIGLIPIKMHENDNAFLANEEKARIRCKRLR
EGIKVSCPEKLOTTLTKYKGNPPIPHSRSEFPAFLKRGANVILLYIQHSSKRIINFL
ALAMTKSATNDNISDTSYLIAMLGNDYIRIEKELDSGVAKDSSSEPSKLEPMGET
PIPAEIHVTLMDPMNVFYPGEVGLVLAEGDIDIGSYQPTVYVAIIVOEVBREAD
NSSFLKIVQIDIGSEYKIVSSIDLKFSRPRESSOSRSDSAPSTSPTEFLTPGLR

```


repeat_region	/note="Alusg repeat: matches 9. .301 of consensus" 38703. .39008
repeat_region	/note="Alusg repeat: matches 1. .306 of consensus" 39790. .40093
repeat_region	/note="Alusx repeat: matches 1. .304 of consensus" 40126. .40416
repeat_region	/note="Alusg repeat: matches 1. .292 of consensus" 40444. .40733
misc_feature	/note="Alusg repeat: matches 1. .292 of consensus" 41322. .41405
repeat_region	/note="Single clone region. Assembly confirmed by restriction digest data." 41541. .41788
repeat_region	/note="Alusg repeat: matches 1. .248 of consensus" 44790. .45101
repeat_region	/note="Alusg repeat: matches 1. .313 of consensus" 45261. .45312
repeat_region	/note="13 copies 4 mer tgtt 88% conserved" 45899. .46206
repeat_region	/note="Alut repeat: matches 1. .307 of consensus" 46754. .47052
repeat_region	/note="Aluv repeat: matches 1. .298 of consensus" 47067. .47365
repeat_region	/note="Aluv repeat: matches 1. .299 of consensus" 47477. .47873
repeat_region	/note="L1M10 repeat: matches 5950. .6322 of consensus" 47889. .48229
repeat_region	/note="Alusx repeat: matches 1. .312 of consensus" 49168. .49212
repeat_region	/note="Alu repeat: matches 85. .126 of consensus" 49620. .49693
repeat_region	/note="L2 repeat: matches 1685. .1757 of consensus" 50704. .51032
misc_feature	/note="Alusg repeat: matches 1. .308 of consensus" 52204. .53009
repeat_region	/note="Cpg island" /evidence=not_experimental 53978. .54137
repeat_region	/note="Alusx repeat: matches 1. .160 of consensus" 54179. .54511
repeat_region	/note="L1M6 repeat: matches 5822. .6172 of consensus" 55665. .55949
repeat_region	/note="Alusx repeat: matches 1. .305 of consensus" 57331. .57390
repeat_region	/note="30 copies 2 mer ga 75% conserved" 57357. .57392
repeat_region	/note="9 copies 4 mer gaga 91% conserved" 57753. .57930
repeat_region	/note="MIR repeat: matches 82. .262 of consensus" 58260. .58389
repeat_region	/note="MIR repeat: matches 2. .153 of consensus" 58564. .58611
repeat_region	/note="24 copies 2 mer ca 93% conserved" 59350. .59533
repeat_region	/note="Alusg repeat: matches 129. .313 of consensus" 59992. .60223
repeat_region	/note="AluDo repeat: matches 85. .299 of consensus" 61036. .61144
repeat_region	/note="L2 repeat: matches 2581. .2696 of consensus" 62188. .62316
repeat_region	/note="TIGER1 repeat: matches 2238. .2418 of consensus" 62008. .62187
repeat_region	/note="Aluv repeat: matches 1. .129 of consensus" 62330. .62363
repeat_region	/note="Alu repeat: matches 261. .294 of consensus" 62362. .62565
repeat_region	/note="TIGER1 repeat: matches 1586. .1787 of consensus" 62566. .62865
repeat_region	/note="Alusg repeat: matches 1. .302 of consensus" 62866. .64385
repeat_region	/note="TIGER1 repeat: matches 46. .1586 of consensus" 64386. .64694
repeat_region	/note="Aluv repeat: matches 1. .306 of consensus"

	repeat_L_region	64695..64713	/note="TIGER1 repeat: matches 29..46 of consensus"
	repeat_L_region	65068..65395	/note="LIPB2 repeat: matches 5405..5733 of consensus"
	repeat_L_region	65396..65569	/note="AluY repeat: matches 136..309 of consensus"
	repeat_L_region	65571..65640	/note="LIPB2 repeat: matches 5728..5791 of consensus"
	repeat_L_region	65696..65717	/note="11 copies 2 mer ta 100% conserved"
	repeat_L_region	65725..66096	/note="LIPB2 repeat: matches 5789..6155 of consensus"
	repeat_L_region	66371..66410	/note="10 copies 4 mer tgtg 82% conserved"
	repeat_L_region	67586..67886	/note="AluJb repeat: matches 1..299 of consensus"
	repeat_L_region	69748..69930	/note="MIR repeat: matches 6..248 of consensus"
	repeat_L_region	70957..71267	/note="AluY repeat: matches 1..311 of consensus"
	repeat_L_region	71279..71413	/note="MER1B repeat: matches 548..680 of consensus"
	repeat_L_region	71411..71737	/note="MER1A repeat: matches 47..485 of consensus"
	repeat_L_region	71780..72075	/note="AlusX repeat: matches 1..295 of consensus"
	repeat_L_region	72145..72256	/note="MER11-internal repeat: matches 42..175 of consensus"
	repeat_L_region	72454..72865	/note="MER11-internal repeat: matches 332..739 of consensus"
	repeat_L_region	72873..73249	/note="MER11-internal repeat: matches 883..1261 of consensus"
	Query Match	100.0%; Score 500; DB 9; Length 92693;	
	Best Local Similarity	100.0%; Pred. No. 5.1e-101;	
	Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
OY	1 atgaattacagaagaccatgcttactcagctcagttcttaataccagaacgttcgaacg	60	
Db	18909 ATGATTTCAGGAGAACCATGTACTACCTCAGCTTTTAATCCAGACATTTGCACG	18850	
OY	61 tcttatcaaggaaagtlaaetlacatctctggcctggcgcagagaatgatgttcaaigtla	120	
Db	18849 TCTTATCAAGGAAGTAATGAATACATCTCGGCCCTGGCAGAAATGATTGTCATCGTGA	18790	
OY	121 tccatttgatgaaaacagaataatcccacatcgtttcacatggcctaagaatggtttgaaaaa	180	
Db	18789 TCCATTATGATAAAAACAANAACACCCTATCTGTTTCATATGCTTAAGATGGTTTGAAAAA	18730	
OY	181 tcttatatacatcttttagagagatttaccttatcttatgtatgagatgacactatccccag	240	
Db	18729 TCTTATATATACATTTTTTAGAGAGATTTCATCTTATTTGATGATGATGCCATTAATCCCCAG	18670	
OY	241 aaactaactagaagaagltcagacalgtctggaactcatactagactcagatgccalcglt	300	
Db	18669 AACCTACTAGAGGAAGGTCAAGACATGTGTGGAACCTATTAGACTCAGAGATTCATCGTT	18610	
OY	301 agtcaatttagacgatgaatctgaaagcacagcttcagaaatctttaagcagacatgttaca	360	
Db	18609 AGTCATTTTAGCGATGTGATCTGAAGGCACAGCTTCCAANAATTTTGAAGACGACATTGTACA	18550	
OY	361 aaaacttggaggggtttgccttaaaaaaatgtagtcatctatatacaaacatccgcttattaa	420	
Db	18549 AAAAATTTGGAGGGTGTGCTTTAAAAAATGTAGATGCATCTATACAAACATCCGCTATTATA	18490	
OY	421 aaaatatatcatcaccatcaccagaatgctgtttcttgagataaagtgaagatgccatt	480	
Db	18489 AAAATATTTTCATTCACCATTAACCAAGTGCCTGTTTTCGAGAGATAATGGAAGAATGCCATT	18430	
OY	481 gcagaaattcgttaatcaaa 500		


```

misc_feature      10026..11426
                    /note="assembly_name:Contig42"
misc_feature      11527..13266
                    /note="assembly_name:Contig44"
misc_feature      13367..14794
                    /note="assembly_name:Contig47"
misc_feature      14895..16054
                    /note="assembly_name:Contig48"
misc_feature      16155..17395
                    /note="assembly_name:Contig49"
misc_feature      17496..19287
                    /note="assembly_name:Contig50"
misc_feature      19388..21294
                    /note="assembly_name:Contig51"
misc_feature      21395..22944
                    /note="assembly_name:Contig52"
misc_feature      23045..24421
                    /note="assembly_name:Contig53"
misc_feature      24522..25870
                    /note="assembly_name:Contig54"
misc_feature      25971..27230
                    /note="assembly_name:Contig55"
misc_feature      27331..28778
                    /note="assembly_name:Contig57"
misc_feature      28879..30893
                    /note="assembly_name:Contig58"
misc_feature      30994..32460
                    /note="assembly_name:Contig59"
misc_feature      32561..33984
                    /note="assembly_name:Contig60"
misc_feature      34085..35285
                    /note="assembly_name:Contig61"
misc_feature      35386..37184
                    /note="assembly_name:Contig62"
misc_feature      37285..39172
                    /note="assembly_name:Contig63"
misc_feature      39273..40874
                    /note="assembly_name:Contig64"
misc_feature      40975..42893
                    /note="assembly_name:Contig65"
misc_feature      42994..44384
                    /note="assembly_name:Contig66"
misc_feature      44485..45999
                    /note="assembly_name:Contig67"
misc_feature      46100..48669
                    /note="assembly_name:Contig68"
misc_feature      48770..50798
                    /note="assembly_name:Contig69"
misc_feature      50899..52809
                    /note="assembly_name:Contig70"
misc_feature      52910..55127
                    /note="assembly_name:Contig71"
misc_feature      55228..58087
                    /note="assembly_name:Contig72"
misc_feature      58188..61004
                    /note="assembly_name:Contig73"
misc_feature      61105..64185
                    /note="assembly_name:Contig74"
misc_feature      64286..67105
                    /note="assembly_name:Contig74"

```

Query Match 99.8%; Score 499; DB 2; Length 99819;
 Best Local Similarity 100.0%; Pred. No. 8.3e-101;
 Matches 499; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 atgattacaggaagcctgctacgtcagcttcaatccagaacgattgcaag 60
    |||
Db 48099 ATGATTACAGGAAGCCATGCTCAGCTTCAATCCAGAAAGATTGTCAGG 48158
    |||
QY 61 tcttaacaagaagtaataacattcgcctgcagagaattgattcaatgta 120
    |||
Db 48159 TCTTAACAAGAAATTAATATTCATTCTGGCCTGCACAGAAATTATTTCAATGGTA 48218
    |||
QY 121 tccatttgataaacaagaatccacatctgttcatggttaagatggttgaaaaa 180
    |||

```

```

Db 48219 TCCATTGTGTAAGAAACAGAAACACCACCTGCTTTCATGCTTAAGATGTTGGAAAA 48278
    |||
QY 181 tcttaataacatttttcagagagattgacttatttgatgagatgccattatcccg 240
    |||
Db 48279 TCTTAATATACATTTTTCAGAGAGATTGACTTATTGATGAGATGCCACTTATVCCCGAG 48338
    |||
QY 241 aactaatacagaagaagtcagacatgltggaactactaatacactagaatccatcgt 300
    |||
Db 48339 AACTATATACAGAGAGATGTCAGACATGTCGAAACCATTTAGCTCAGATTCATCGTT 48398
    |||
QY 301 agtcaattagaagatgaatctggaagcacagcttccagaattttagaagacatgtaca 360
    |||
Db 48399 AGTCATTTAGACGATGAATCTGACACACGCTCCAGATTTTTCAGACATTTGTACA 48458
    |||
QY 361 aaactggaagggttgcttcttaaaaaatgatgatcatatcacacatccgatttaa 420
    |||
Db 48459 AAAACTGGAGGGTGTGCTTAAAAAATGATGATCATATACACATCCCTTATTTAA 48518
    |||
QY 421 aaataatacatcacattacacaaagtgtgttttcagacataatggaagaatgcatt 480
    |||
Db 48519 AAAATATATTCATTCACCATTCACCATTCGCTGTTTGCAGATATGAGAAAGATGCCATT 48578
    |||
QY 481 gcagaaatgtgtaatcaa 499
    |||
Db 48579 GCAGAAATTGTGTAATCAA 48597
    |||

```

RESULT 5
 AC069017 174140 bp DNA linear HTG 12-SEP-2001
 LOCUS Mus musculus clone MGS3-342116, WORKING DRAFT SEQUENCE, 5 unordered
 DEFINITION pieces.
 AC069017
 VERSION AC069017.20 GI:15559167
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE
 AUTHORS

1 (bases 1 to 174140)
 Metzger, M.L., Lewis, L.R., Hume, J., Edwards, C., Harris, C.,
 Dederich, D., Thomas, S., Okwuonu, G., Carllock, C., Garner, T.,
 Addison, S., Pace, A., Williams, G., Bonin, D., Brooks, A., Brown, J.,
 Buhay, C., Bunac, C., Burkett, C., Chacko, D., Chen, G., Chen, Z.,
 Cox, C., Davis, C., Delgado, O., Ding, Y., Dugan-Rocha, S.,
 Fernandez, C., Ferraguto, D., Forcum-Tansey, J., Gill, R.,
 Gorrell, J.H., Gunaratne, P., Haller, G., Hernandez, J., Hognes, M.,
 Hosak, H., Hou, X., Huber, J., Jackson, L., Jia, Y., Kelly, J., Kelly, S.,
 Kovar, C., Liu, J., Liu, W., Louisseged, H., Lozado, R.J., Martin, R.,
 Massey, E., McLeod, M.P., Mei, G., Moore, S., Morgan, M., Morris, S.,
 Neal, D., Nelson, A., Nguyen, R., Nguyen, N., Ogun, M., Parish, B.,
 Perez, L., Relter, D., Say, J., Shen, H., Vasquez, L., Watlington, S.,
 Williamson, A., Wrensford, G., Zhou, X., Bouck, J., Hodgson, A.,
 Muzny, D.M., Rives, M., Scherer, S., Sodergren, E., Weinstock, G.,
 Worley, K. and Gibbs, R.

TITLE Direct Submission
 JOURNAL
 REFERENCE
 COMMENT
 2 (bases 1 to 174140)
 Worley, K.C.
 Direct Submission
 Submitted (17-MAY-2000) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Sep 12, 2001 this sequence version replaced gi:14787161.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: MGAO
 Center clone name: MGS3-342116

ACCESSION AE001381 AE001382
VERSION AE001381.1 GI:3845124
KEYWORDS malaria parasite P. falciparum.
SOURCE Plasmodium falciparum
ORGANISM Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE 1 (bases 1 to 12029)
AUTHORS Gardner,M.J., Wetzelin,H., Carucci,D.J., Cummings,L.M., Araavind,L., Koonin,E.V., Shalton,S., Mason,T., Yu,K., Fujii,C., Pederson,J., Shen,K., Jhing,J., Aston,C., Lai,Z., Schwartz,D.C., Perle,M., Salzberg,S., Zhou,L., Sutton,G.G., Clayton,R., White,O., Smith,H.O., Fraser,C.M., Hoffman,S.L. et al.
TITLE Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum
JOURNAL Science 282 (5391), 1126-1132 (1998)
MEDLINE 99021743
REMARK Erratum: [Published erratum appears in Science 1998 Dec 4;282(5395):1827]]
REFERENCE 2 (bases 1 to 12029)
AUTHORS Gardner,M.J.
TITLE Direct Submission
JOURNAL Submitted (02-NOV-1998) The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20814, USA
FEATURES
source
1..12029
/organism="Plasmodium falciparum"
/db_xref="taxon:5833"
/chromosome="2"
/gene="PFB0210c"
/complement(105..1619)
/gene="PFB0210c"
/complement(105..1619)
/note="Identified by sequence similarity; putative"
/codon_start=1
/product="monosaccharide transporter"
/protein_id="A6C71831.1"
/db_xref="GI:3845125"
/translation="MTRSSKIDICSENEKKNGSGFSSFRVLSACIASIFFGV
SVLNTKNTIVEFEMCKGKDRNLCSNNTIOSSFLASVIGAVJGCGSGVLVPG
RLSLIILINPEFLVSLISITHFPTILFARLSEGIQVLVSPVMISEMTKDK
KGAVGMHOLFIFGIFVAVMLGAMGEGPKADSTPLTSFAKIMRLMFLPSSVL
IGIALVVEFEETPYEFLEKRIEESKNILKKIYEDNVEDPLNKEAVEONESAK
KNSLSLSALKIPSYRYVILIGLGLGLOOFGINLVNSNELYKPELDSHLITLS
VYVTAVPLMTPAIYIVIEKIGKTLINGCVCVLAIVPTAIANEINRNSNVKILS
IVATEPMIISFAVSGIPVMTIYLHEMPSEIKDSASLASLVNVCALIVFSDIIL
KPSILFLIVFSVMSILTEFFIFEFIKETKGEGIGSPYITMEEROKHMTKSVV"
/complement(3799..5049)
/gene="PFB0215c"
/complement(3799..5049)
/gene="PFB0215c"
/note="Identified by sequence similarity; putative"
/codon_start=1
/product="protein with Egl-1like 3'-5' exonucl. domain"
/protein_id="A6C71832.1"
/db_xref="GI:3845126"
/translation="MTRCKLNINIMKRRVRRNSYCYNNININEKIHKFDNDKRN
IKTISDCKSCRECVDEIKNGNTNLLKDFNMKMGIDEGIKGKIGVSTIQCYEDI
YLPDIYKCNVYLFYINIKIDLECDIIVTHHCRCDSLNYQVNIHLNILDQVA
VYDLKNNNNYVTFYOISYDILKYLFLVNNHRIEFHKMILDNYILKRPIMKEIL
SYAIDVYILKPLICILDKFEIKOKKEOEKNKYVNDQNNKIKOEKDKSNLLO
SKNISSPNNODLYHFKETIIDIILSHSKYVNOPIFNHSHKDEKELOKMGILGMYV
CNRTKMYILKLMKRGVILINYNOKYKIEIGDIVAIVNFRNDYINGLIDEKILLD
AOKYIPREVLONIQFLNEEKI"
/complement(5564..6843)
/gene="PFB0220w"
/join(5564..5681,5897..6843)
/note="PFB0220w"
/note="Identified by sequence similarity; putative"
/codon_start=1
/product="ubiquitinone biosynthesis methyltrans."
/protein_id="A6C71833.1"
/db_xref="GI:3845127"
/translation="MKITKINYNMFKITVWVGIIISKASHSFSTQHGQYDKSRICNF

GFQVSEETKSRILVNLFSNVCKKDYIMDMMSLVLHREKQDFVKELDILKYSHYN
IDQVYQHKDHSNNEKTOKKNEBNTSDNGYNNNSVSYSDIPNYKILDLAGCTGDLA
FRLERKSKFYLKKNOSIPDHLSTQOYLPHIIVCNDMLNVGKKAATLQDONL
TWLVQNAENLESVESNSIDVTLISFGIIRNETNPQALKEHRLKGGRLCLEFSKY
OCHIFNIFYKFLNANVPIIGKVANDMAKYVASIESIQFETLPDELAOLFHOANFN
ITYTMTMGIVSHSAVKLY"
/complement(7198..8724)
/gene="PFB0225c"
/complement(7198..8724)
/gene="PFB0225c"
/note="Predicted by Glimmer"
/codon_start=1
/product="hypothetical protein"
/protein_id="A6C71834.1"
/db_xref="GI:3845128"
/translation="MGLPRKKNVPCYGYDRIHSLELRINNIILKYSNIETCKN
KKCLDELVLASYDNFLKKEKYEYCKLDGYINEDEKIKINEVKGGRKKIDCTPN
NNKIFPVNHLINDDLFKRKKKKKKKTKLKNCKNDRNLKNEKDDHVES
YTKKKMLNNNNNNNITNMVSEFGYNKRRKYTNKCIINEDENNKMDENOHDKRN
INTNLHDDKNDKIRKHSITQFLYSNNKEKIIISKNVLKDESTNITKRNKVDN
IDHOKHREKREKKSLENMISIIIEYQSDICNVLEENKEEKYNNLQKDVITN
CNDKVKLEEYHHEKELNVQIINDDKNEKKNNKKKKKKKKKKKKKKKKKKKKKK
EKKNNKKEKK
DYELLYEEKIINVTKKDELSKKNKSEKDIYDILSEYNTLQYTSFLDYKKNME

BASE COUNT 4304 a 1428 c 922 g 5375 t
ORIGIN
Query Match 8.4%; Score 42.2; DB 3; Length 12029;
Best Local Similarity 49.3%; Pred. No. 9.8;
Matches 110; Conservative 0; Mismatches 113; Indels 0; Gaps 0;
OY 33 agcttcaaatccagaagatttcagcgttcatacaaggaaagtataataatcctggc 92
DB 4360 AGGTTTAATAATATATTCATCTGGATGACATATGAATAATTATCTTTCATATAAGGCT 4419
OY 93 ctggcagaagaattgatgttcacatgcatcattgttgaaacagaatcacccatctg 152
DB 4420 CTTTAATAATAATAATATCTAGTGATCTTTATGAAATAATTTTATGATATAT 4479
OY 153 ttcaatgcttaagaatgcttggaataatcttataataatcatttcagaggaattgact 212
DB 4480 ATTATATAAATAATATTTTATATAATCATCTACTAATTTTGAAAGATATTTGTATA 4539
OY 213 tattgatgagatgccctataccccgaactatacagagga 255
DB 4540 ATTATATATTATTTTATATAATAATTATATGCTTGTGTA 4582
RESULT 9
AC096291/c

LOCUS AC096291 116127 bp DNA linear HMG 20-DEC-2001
 DEFINITION Rattus norvegicus clone CH230-38P14, *** SEQUENCING IN PROGRESS
 ACCESSION AC096291
 VERSION AC096291.2 GI:17943988
 KEYWORDS HMG; HMGs_PHASE1.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Rattus.
 1 (bases 1 to 116127)
 Muzny,D.M., Adams,C., Adio-Oduola,B., All-osman,F.R., Allen,C.,
 Alstbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbarella,J.,
 Benton,J., Blincoe,K., Blankenburg,K., Bonin,D., Bouck,T.,
 Bowls,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
 Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
 Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
 Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
 Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
 Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
 Dem,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
 Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
 Ellaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
 Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
 Garza,N., Gill,R., Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S.,
 Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hayes,A.,
 Hernandez,J., Hernandez,O., Hodgson,A., Hognes,M., Holloway,C.,
 Hollins,B., Homs,F., Howard,S., Huber,J., Hulys,S., Hume,J.,
 Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jollivet,S.,
 Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
 Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
 Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
 Louised,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
 Ma,J., Maheshwari,M., Mapa,P., Martin,R., Martindale,A.,
 Martinez,E., Massey,E., Mawhinney,E., McLeod,M.P., Meador,M.,
 Mel,G., Metzger,M., Miner,G., Miner,Z., Mitchell,T., Mohabab,K.,
 Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
 Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
 Ogutu,N., Okunolu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
 Peery,J., Perez,L., Peters,L., Pickens,R., Prims,E., Pu,L.L.,
 Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojudoan,I., Rolfe,M.,
 Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshari,N.,
 Sisson,I., Sodergren,E., Sonalke,T., Sparks,A., Stanley,H.,
 Stone,H., Sutton,A., Svatek,A., Taber,P., Tamerisa,A., Tamerisa,K.,
 Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
 Thomas,S., Usmani,K., Vasquez,L., Verra,V., Villalob,D., Vinson,R.,
 Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
 Watlington,S., Williams,G., Williamson,A., Wleczka,R., Wooden,S.,
 Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zortilla,S., Nelson,D.,
 Weinstein,G. and Gibbs,R.
 Direct Submission
 Unpublished
 2 (bases 1 to 116127)
 Worley,K.C.
 Direct Submission
 Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Dec 20, 2001 this sequence version replaced gi:15627911.
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GEWI
 Center clone name: CH230-38P14
 ----- Summary Statistics
 Assembly program: Phrap; version 0.990329Pfirst call to
 findphraplist
 Consensus quality: 83123 bases at least Q40
 Consensus quality: 91227 bases at least Q30

Consensus quality: 96818 bases at least Q20
 Estimated insert size: 71455; sum-of-contigs estimation
 Quality coverage: 0x in Q20 bases; aabase-fp estimation
 Quality coverage: 0.9x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 61 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 *
 * 1 3916: contig of 3916 bp in length
 * 3917 4016: gap of unknown length
 * 4017 7194: contig of 3178 bp in length
 * 7195 7294: gap of unknown length
 * 7295 11673: contig of 4379 bp in length
 * 11674 11773: gap of unknown length
 * 11774 14734: contig of 2961 bp in length
 * 14735 14834: gap of unknown length
 * 14835 18012: contig of 3178 bp in length
 * 18013 18112: gap of unknown length
 * 18113 19884: contig of 1772 bp in length
 * 19885 19984: gap of unknown length
 * 19985 22062: contig of 2078 bp in length
 * 22063 22162: gap of unknown length
 * 22163 25052: contig of 2850 bp in length
 * 25053 25153: gap of unknown length
 * 25153 27939: contig of 2787 bp in length
 * 27940 30903: gap of unknown length
 * 30904 31003: gap of unknown length
 * 31004 33492: contig of 2469 bp in length
 * 33493 33592: gap of unknown length
 * 33593 36532: contig of 2940 bp in length
 * 36533 38663: gap of unknown length
 * 38663 38963: contig of 2231 bp in length
 * 38964 42247: gap of unknown length
 * 42248 42347: gap of unknown length
 * 42348 43940: contig of 1493 bp in length
 * 43941 47252: gap of unknown length
 * 47253 47352: contig of 3312 bp in length
 * 47353 50104: gap of unknown length
 * 50105 50204: gap of unknown length
 * 50205 52302: contig of 2098 bp in length
 * 52303 52402: gap of unknown length
 * 52403 54427: contig of 1925 bp in length
 * 54428 54427: gap of unknown length
 * 54428 55733: contig of 1306 bp in length
 * 55734 55833: gap of unknown length
 * 55834 57172: contig of 1339 bp in length
 * 57173 58476: gap of unknown length
 * 58477 58576: contig of 1204 bp in length
 * 58577 60321: gap of unknown length
 * 60322 60421: contig of 1745 bp in length
 * 60422 62141: gap of unknown length
 * 62142 62241: gap of unknown length
 * 62242 63722: contig of 1481 bp in length
 * 63723 63822: gap of unknown length
 * 63823 65055: contig of 1233 bp in length
 * 65056 65155: gap of unknown length
 * 65156 66343: contig of 1188 bp in length
 * 66344 66444: gap of unknown length
 * 66444 67996: contig of 1453 bp in length
 * 67997 69108: gap of unknown length
 * 69109 69208: contig of 1112 bp in length
 * 69209 69208: gap of unknown length


```

* 69209 70970: contig of 1762 bp in length
* 70971 71070: gap of unknown length
* 71071 72523: contig of 1453 bp in length
* 72524 72623: gap of unknown length
* 72624 74933: contig of 2330 bp in length
* 74934 75053: gap of unknown length
* 75054 76089: contig of 1036 bp in length
* 76090 76189: gap of unknown length
* 76190 77687: contig of 1498 bp in length
* 77688 77788: gap of unknown length
* 77788 79293: contig of 1506 bp in length
* 79294 79393: gap of unknown length
* 79394 80715: contig of 1322 bp in length
* 80716 80815: gap of unknown length
* 80816 82192: contig of 1377 bp in length
* 82193 82292: gap of unknown length
* 82293 83772: contig of 1480 bp in length
* 83773 83872: gap of unknown length
* 83873 84994: contig of 1122 bp in length
* 84995 85094: gap of unknown length
* 85095 86566: contig of 1562 bp in length
* 86567 86756: gap of unknown length
* 86757 88212: contig of 1456 bp in length
* 88213 88312: gap of unknown length
* 88313 89343: contig of 1031 bp in length
* 89344 89443: gap of unknown length
* 89444 90867: contig of 1424 bp in length
* 90868 90967: gap of unknown length
* 90968 92392: contig of 1425 bp in length
* 92393 92492: gap of unknown length
* 92493 94094: contig of 1602 bp in length
* 94095 94194: gap of unknown length
* 94195 95340: contig of 1146 bp in length
* 95341 95440: gap of unknown length
* 95441 97148: contig of 1708 bp in length
* 97149 97248: gap of unknown length
* 97249 98441: contig of 1193 bp in length
* 98442 98541: gap of unknown length
* 98542 99693: contig of 1152 bp in length
* 99694 99793: gap of unknown length
* 99794 101143: contig of 1350 bp in length
* 101144 102563: contig of 1320 bp in length
* 102564 102663: gap of unknown length
* 102664 103974: contig of 1311 bp in length
* 103975 104074: gap of unknown length
* 104075 105257: contig of 1183 bp in length
* 105258 105357: gap of unknown length
* 105358 106636: contig of 1279 bp in length
* 106637 106736: gap of unknown length
* 106737 107801: contig of 1065 bp in length
* 107802 107901: gap of unknown length
* 107902 109084: contig of 1183 bp in length
* 109085 109184: gap of unknown length

```

Query Match 8 4%; Score 42; DB 2; Length 116127;
 Best Local Similarity 50.0%; Pred. No. 7.6; Indels 0; Gaps 0;
 Matches 105; Conservative 0; Mismatches 105;

```

QY 255 aagtcagacatgctgtgaacatcattgacgagatcattcattcgttattgattgagc 314
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 75355 AAGATCATTAAGTATGTCGCAAAATAGAAATTATAGAGTTTATATTAATAGATGACCA 75296
QY 315 atgaatctgaagcagcgtccagaaatttttagcacacattgttacaanaacttgaggg 374
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 75295 TTAATAAATGATTAACATTAATCCATTTTAAGAATTCAATAGAAAGAAAAATGCAATTG 75236
QY 375 ttgtccttaaaaattagatgcatctatacaacatccgccttataaanaatatattcatt 434
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 75235 TGAATTAATAGTTTAATTAATTAATCAACTACTATTATTCAAAATAAAGTTTAGCAAGA 75176
QY 435 caccattaccaagtcgtctttagcatataa 464
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Db 75175 TACAGAAAGAAATTTCTGACTTCACATATA 75146

RESULT 10
 CEY39A1A 110960 bp DNA linear INV 11-DEC-2001
 Caenorhabditis elegans cosmid Y39A1A, complete sequence.

LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 REMARK
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL

1 (sites)
 none.
 Genome sequence of the nematode *C. elegans*: a platform for investigating biology. The *C. elegans* Sequencing Consortium
 Science 282 (5396), 2012-2018 (1998)
 99069613
 The *C. elegans* Sequencing Consortium.
 2 (bases 1 to 110960)
 Wall, M.
 Direct Submission
 Submitted (22-SEP-1998) Nematode Sequencing Project, Sanger
 Institute, Hinxton, Cambridge CB10 1SH, England and Department of
 Genetics, Washington University, St. Louis, MO 63110, USA. E-mail:
 jesus@anger.ac.uk or rwn@nematoide.wustl.edu
 Coding sequences below are predicted from computer analysis, using
 predictions from GeneFINDER (P. Green, U. Washington), and other
 available information.
 Current sequence finishing criteria for the *C. elegans* genome
 sequencing consortium are that all bases are either sequenced
 unambiguously on both strands, or on a single strand with both a
 dye primer and dye terminator reaction, from distinct subclones.
 Exceptions are indicated by an explicit note.
 IMPORTANT: This sequence is not the entire insert of clone Y39A1A.
 It may be shorter because we only sequence overlapping sections
 once, or longer because we arrange for a small overlap between
 neighbouring submissions.
 The true left end of clone Y39A1 is at 1 in this sequence. The true
 left end of clone W09D10 is at 110854 in this sequence. The start
 of this sequence (1..106) overlaps with the end of sequence
 AL032621.
 The end of this sequence (110854..110960) overlaps with the start
 of sequence Z93785.
 For a graphical representation of this sequence and its analysis
 see: [http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?](http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?name=Y39A1A)
 name=Y39A1A
 IMPORTANT: This sequence is NOT necessarily the entire insert of
 the specified clone. It may be shorter because we only sequence
 overlapping sections once, or longer because we arrange for a small
 overlap between neighbouring submissions.
 Location/Qualifiers

1..110960
 /organism="Caenorhabditis elegans"
 /db_xref="taxon:6239"
 /chromosome="III"
 /clone="Y39A1A"
 complement(join(1078..1247,2168..2941,3412..3482,
 3561..3777,4206..4296))
 /gene="Y39A1A.1"
 complement(join(1078..1247,2168..2941,3412..3482,
 3561..3777,4206..4296))
 /gene="Y39A1A.1"
 /note="cDNA EST EMBL:700867 comes from this gene"
 /codon_start=1
 /protein_id="CAA21019.2"
 /db_xref="GI:15718274"
 /db_xref="SPTREMBL:O9XX21"
 /translation="MSKKEEITFEFRLEKENRPESSKKEEDENSTEEEMTLNHNASYTLD
 HSAPFIADKDFKYNOLNLFHFRMYKDYKGGSSRIYVNIFFLSKLVKIKFCGLK
 OGVLSTAKINFTYHNTFTWPLVKIKNGPVRLVLPDQNSRRITIIYSALAGRPQNN

gene
CDS
complement(join(7136..7306,7430..7540,7620..7670))
/gene="y39A1A.2"
/codon_start=1
/protein_id="CAA21018.1"
/db_xref="GI:380845"
/db_xref="SPTREMBL:O9XK22"
/translation="MSSEGSYPIKLADRFLHYIDFNADPYTCISASSGTVSYDVSYS
RAYPPNDALILCCQOPPPEPKSTIKTYVRKKELMEIPVDIEFMRAIOLKVKDGRRE
NSATGNOR"

gene
CDS
join(18497..8559,9470..9553,9733..9848,10170..10270,
10365..10416,10699..10768)
/gene="y39A1A.4"
/join(8497..8559,9470..9553,9733..9848,10170..10270,
10365..10416,10699..10768)
/gene="y39A1A.4"
/note="predicted using GeneFinder
containing similarity to Pfam domain: PF01754 (A20-like zinc
finger). Score=35.7, E-value=3.5e-07, N=1"

gene
CDS
/codon_start=1
/protein_id="CAA21016.1"
/db_xref="GI:380843"
/db_xref="SPTREMBL:O9XX24"
/translation="MEIRPFOEIFEVLFTLPDPCFOPPHKVMSPDKHGEEDPKAKMHL
KRANDLCVNGCGFCGYCTROMENRNCSCKAHONKECODPAKNISLSITPOERKRS
TTESRSRGIKNLFKTSPIPEGSTSSPTSTSASTPTRAEAVCFRTKTNFSAISSFRL
P"

gene
CDS
join(14209..14677,14977..15126,15303..15440,15752..16004,
16352..16433,16754..16939,18701..19006)
/gene="y39A1A.5a"
/join(14209..14677,14977..15126,15303..15440,15752..16004,
16352..16433,16754..16939,18701..19006)
/gene="y39A1A.5a"
/note="contains similarity to pfam domain: PF02204
(Vacuolar sorting proteoin 9 (VPS9) domain), score=36.7,
E-value=1.7e-07, N=1"

gene
CDS
CDNA EST yk210c11.3 comes from this gene
CDNA EST yk210c11.5 comes from this gene
CDNA EST yk532c5.5 comes from this gene
CDNA EST yk556b4.5 comes from this gene
/codon_start=1
/protein_id="CAA21021.1"
/db_xref="GI:380848"
/db_xref="SPTREMBL:O9XX19"
/translation="MAOEIARSYVKAIVNRKISEMRSSDDMSKELYWVSYOYLGERIGH
SLFDSPDCKVEDWDQVEKYISIFCSYIFCAHHEEVDAISLODRIRLHWTAG
PLETWAVRKQVVRKDIDEAISLEIAKASAPFKLDCLSKSAIPEALKESASNT
ISYKPGLIWMNNKMKLMKKRNENOKIRKRSKIRIFYEFKMYPLTIRAKTSADDFLT
LTIYLRKSGPLIOSNVAFISRFALIPARLMGGAAYFTNLSCALEFAARNNHESLDM
EKSEAEATYSGHLLAPPLVISINACNOALTYLEGTEITLYNAKAGSLANKNLTMHGK
SDDDLEMLAVKIKEYVDYPJDDEYMNMOQSIFAEKCEADLIVLSIRSSESGGTUT
DROPTPDPTSIESPLIAEIOAILNLTSSGNOHBEGDELGAAGPLDEMELVALYKIS
TFKKAAADNTVASAKRSGSTQSPVITAASTQORASPALAOYEKESVNAQPREPRSRSK
HHAKRRKMDSDALKASAON"

gene
CDS
join(18036..18110,18701..19006)
/gene="y39A1A.5b"
/join(18036..18110,18701..19006)
/gene="y39A1A.5b"
/note="CDNA EST yk112f11.3 comes from this gene
CDNA EST yk112f11.5 comes from this gene
CDNA EST yk68hl.3 comes from this gene
CDNA EST yk68hl.5 comes from this gene
CDNA EST yk451e4.3 comes from this gene
CDNA EST yk437d12.3 comes from this gene
CDNA EST yk437d12.5 comes from this gene
CDNA EST yk532c5.3 comes from this gene
CDNA EST yk540g3.3 comes from this gene

gene
 CDS
 cdna EST yk541e1.3 comes from this gene
 cdna EST yk556b4.3 comes from this gene
 cdna EST yk540g3.5 comes from this gene
 cdna EST yk541e1.5 comes from this gene
 /codon_start=1
 /protein_id="CAA21020.1"
 /db_xref="GI:3880847"
 /db_xref="SPRTEMBL:Q9XX20"
 /translation="MLELLEALLGLVAVALLTWHRKEVEKGEGGLGAGPILEMEALV
 LKISTIFKQAGAGNTVTSASKRSTGSPVIRFAEASTONASPAIAQVEKFESVROPREP
 RSRKSHAKHRKSKMDSALKAASON"
 join(19588, 19881, 20463, .20639, 20856, .20982, 21728, .21912)
 /gene="Y39A1A.6"
 join(19588, 19881, 20463, .20639, 20856, .20982, 21728, .21912)
 /gene="Y39A1A.6"
 /note="cdna EST yk471b2.3 comes from this gene
 cdna EST yk471b2.5 comes from this gene"
 /codon_start=1
 /protein_id="CAA21022.1"
 /db_xref="GI:3880849"
 /db_xref="SPRTEMBL:Q9XX18"
 /translation="MLRVLVRTGSEACRAVSTRAVLVHPVAKSEDLISAKIEKDRS
 KLRVPLQKDEOLSKSYFAPEMDLKKKPNLDEGFVPLKGYMTPEKMEYKNVVP
 PNYIVPENGPLPKPEKVEFKESVHSPKRMAGOLVAKVMVDEAITOLDQOLKACN
 LLMPTIKAKSRADDERHIEIPESOMYADAPVOSNLYKGGARRAHNMNITRRYIH
 IFLVEEGPAPQOKORHPQKGMWHDYVLYLSRYKYSI"
 complement(join(22174, .22353, 23082, .23416, 24281, .24329))
 /gene="Y39A1A.7"
 complement(join(22174, .22353, 23082, .23416, 24281, .24329))
 /gene="Y39A1A.7"
 /note="predicted using GeneFINDER"
 /codon_start=1
 /protein_id="CAA21015.1"
 /db_xref="GI:3880842"
 /db_xref="SPRTEMBL:Q9XX25"
 /translation="MTSITSFLLILLFVAACHSVPCDGCCELSHPDVPICDGV
 GLNRPPLPHSPPLGPHFLATCNDIDITPIASIMKASFPLQDITDQGNRLNCSDL
 VHLKRLPVLSDCNEKPEFLQCDNLDKNDMKCRFLTLKEMWAFKDLVNNKAKEMWA
 BETLEAKSWFSAPFKKRIALIGLSD"
 join(25750, .25789, 25839, .25909, 25963, .26088, 26137, .26405,
 27339, .27744, 28541, .28596, 28650, .28832, 28993, .29064,
 29500, .29707, 30169, .30268, 30345, .30355)
 /gene="Y39A1A.23"
 join(25750, .25789, 25839, .25909, 25963, .26088, 26137, .26405,
 27339, .27744, 28541, .28596, 28650, .28832, 28993, .29064,
 29500, .29707, 30169, .30268, 30345, .30355)
 /note="cdna EST EMBL:AU112975 comes from this gene"
 cdna EST EMBL:AU112975 comes from this gene"
 /codon_start=1
 /protein_id="CAC42372.1"
 /db_xref="GI:14530622"
 /translation="MKGAOFVOSNIKIMRSIALSKISPDVLIENVSEGLFFKTYN
 RSKCVRFAPEFNNACDYSIMINKKAYNICLSMKSAORIFKGYAFEGKNVCEPRT
 DPKRRMKNVLOMNYGKRTISCRIFEKLAESFIDERTTHAKIREGSMILHKTYPNSGC
 RNIVRVPASTLLPIFVOKGDIENVTKMTVDGLIIRNFHSLDVTNMVGEKAKKY
 KTEITITCEKLTFRKIKIPIVPEFSIKESILVIFADOLGSEVMYDLPKPIIVSI
 EAHNFDEIALATMGSDIEDLDGLIKETMAHEEEDKSTAHSSSRKSKAIDT
 SSGTOKSKKSESLSQETTRRSQSLSPRNRFVEIPVADGSMWDRQVYEQNEPSP

Db 17474 CCAAAACATCAATATAGTTTCACCTGGATTTTGGAAATTTGGCGATTTT 17415
 QY 345 tagcagacattgtacaaactcggagggttgccttaaaaattagatgc 396
 Db 17414 TTCTAGAAATTTTAAATAATTTTTCGAATTAATAAATCCGCTAC 17363

RESULT 11

LOCUS AC060835 160624 bp DNA linear PRI 09-JAN-2002
 DEFINITION Homo sapiens BAC clone RP11-785F11 from 4, complete sequence.
 AC060835
 VERSION AC060835.7 GI:15145631
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE 1
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 TITLE Sulston, J.E. and Waterston, R.
 JOURNAL Toward a complete human genome sequence
 MEDLINE Genome Res. 8 (11), 1097-1108 (1998)
 REFERENCE 2 (bases 1 to 160624)
 99063792
 AUTHORS Tomlinson, C., Abbott, A., Hawkins, M., Dixon, R. and Boyer, E.
 TITLE The sequence of Homo sapiens BAC clone RP11-785F11
 JOURNAL Unpublished (2001)
 REFERENCE 3 (bases 1 to 160624)
 3
 AUTHORS Waterston, R.H.
 TITLE Direct Submission
 JOURNAL Submitted (20-APR-2000) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 REFERENCE 4 (bases 1 to 160624)
 4
 AUTHORS Waterston, R.H.
 TITLE Direct Submission
 JOURNAL Submitted (09-AUG-2001) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 REFERENCE 5 (bases 1 to 160624)
 5
 AUTHORS Waterston, R.
 TITLE Direct Submission
 JOURNAL Submitted (09-JAN-2002) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 On Aug 9, 2001 this sequence version replaced gi:14670156.

Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: <http://genome.wustl.edu/gsc>
 Contact: sapiens@wustl.wustl.edu
 Summary Statistics
 Center project name: H_NH0785F11

NOTICE: This sequence may not represent the entire insert of this
 clone. It may be shorter because we only sequence overlapping
 clone sections once, or longer because we provide a small overlap
 between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded, sequenced with an alternate
 chemistry, or covered by high quality data (i.e., phred quality >=
 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by sequence
 from more than one subclone; and the assembly was confirmed by
 restriction digest.

MAPPING INFORMATION:
 Mapping information for this clone was provided by Dr. John D.
 McPherson, Department of Genetics, Washington University, St. Louis
 MO. For additional information about the map position of this
 sequence, see <http://genome.wustl.edu/gsc>

FEATURES

source

SOURCE INFORMATION:
 The RPCT-11 human BAC library was made from the blood of one male
 donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E.,
 Tatem, M., Catanes, J.J. and de Jong, P.J. (1998) An improved
 approach for construction of bacterial artificial chromosome
 libraries. Genomics 51:1-8. The clone may be obtained either from
 Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong
 and coworkers at the Roswell Park Cancer Institute
 (<http://bacpac.med.buffalo.edu>)
 VECTOR: pBAC3.6
 NEIGHBORING SEQUENCE INFORMATION:
 The clone sequenced to the left is RP11-114M4; the clone sequenced
 to the right is RP11-327I3. Actual start of this clone is at base
 position 1 of RP11-785F11; actual end is at base position 160624 of
 RP11-785F11.

Location/Qualifiers

1. 160624
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="4"
 /map="4"
 /clone="RP11-785F11"
 /clone_lib="RPCT-11"
 78. 478
 /rpt_family="L2"
 497. 1140
 /rpt_family="L1"
 3053. 3103
 /rpt_family="AT_rich"
 3390. 3638
 /rpt_family="L2"
 3656. 4189
 /rpt_family="ERV_L"
 4201. 4975
 /rpt_family="L2"
 4983. 5188
 /rpt_family="MER1_type"
 5193. 5243
 /rpt_family="L2"
 5299. 5534
 /rpt_family="L2"
 6010. 6383
 /rpt_family="MALR"
 6471. 7561
 /rpt_family="L1"
 7598. 8936
 /rpt_family="L1"
 8937. 9234
 /rpt_family="Alu"
 9235. 9420
 /rpt_family="L1"
 9388. 9484
 /rpt_family="L1"
 9489. 10100
 /rpt_family="L1"
 10114. 11038
 /rpt_family="L1"
 11039. 11289
 /rpt_family="Alu"
 11290. 13382
 /rpt_family="L1"
 13375. 13395
 /rpt_family="AT_rich"
 13396. 13440
 /rpt_family="(TA)n"
 14049. 14187
 /rpt_family="MER1_type"
 15199. 15223
 /rpt_family="(A)n"
 16128. 16439
 /rpt_family="L1"
 16440. 16743

```
repeat_region /rpt_family="Alu" 16744. 16874
/rpt_family="L1"
repeat_region /rpt_family="ERV1" 16916. 17257
/rpt_family="ERV1" 17413. 17440
repeat_region /rpt_family="AT-rich" 18356. 18659
/rpt_family="Alu" 18631. 18671
repeat_region /rpt_family="(CAAA)n" 18807. 18859
/rpt_family="AT-rich" 18884. 19097
/rpt_family="L1" 19070. 19091
repeat_region /rpt_family="AT-rich" 19110. 19547
/rpt_family="L1"
repeat_region /rpt_family="L1" 21198. 21237
/rpt_family="(T)n" 21214. 21497
repeat_region /rpt_family="Alu" 23018. 23045
/rpt_family="AT-rich" 23028. 23698
repeat_region /rpt_family="L1" 23692. 24220
/rpt_family="L1" 24308. 25255
repeat_region /rpt_family="L1" 30504. 31045
/rpt_family="L1" 31156. 31254
repeat_region /rpt_family="MIR" 31564. 31595
/rpt_family="AT-rich" 31794. 31882
repeat_region /rpt_family="AT-rich" 32101. 32426
/rpt_family="ERV1" 32394. 32414
repeat_region /rpt_family="AT-rich" 32418. 32479
/rpt_family="(TC)n" 32488. 32523
repeat_region /rpt_family="(CA)n" 32613. 32667
/rpt_family="(CAGAGA)n" 33504. 33524
repeat_region /rpt_family="AT-rich" 34257. 34475
/rpt_family="MIR" 34748. 35031
repeat_region /rpt_family="Alu" 35304. 35454
/rpt_family="MIR" 35602. 35667
repeat_region /rpt_family="ERV1" 37934. 39378
/rpt_family="L2" 38783. 38810
repeat_region /rpt_family="(CA)n" 42850. 42890
```

Query Match	8.3%;	Score 41.6;	DB 9;	Length 160624;
Best Local Similarity	48.3%;	Pred. No. 8.8;		
Matches 116;	Conservative	0;	Mismatches 124;	Indels 0;
				Gaps 0;

Oy		catttgatgaagacagaatacccaactgtttcatgcttaagatggttggaaaatc	182
Db	19161	CACCAATTTTATAACAAGATCCCTATGCATTCATTGCGAAGAATAATCGGGAGACAC	19220

19161	CACCTAATTATATACAAAGATCCCTATGCAATTCATGTGMAAGAAATCTGGAGACAC	19220
Quality coverage: 6.0 in Q20 bases; agarose-ef		
Quality coverage: 6.0 in Q20 bases; sum-of-contigs		

[illegible]

RESULT	12				
AC020669/c					
LOCUS					
DEFINITION					
AC020669		168982 bp	DNA	linear	HTG 26-MAY-2000
	Homo sapiens clone RP11-114M4,	WORKING	DRAFT	SEQUENCE	5 unordered
pieces					
AC020669					
AC020669.4	GI:7107781				
KEYWORDS					
HTG; HTGS_PHASE1; HTGS_DRAFT.					
SOURCE	human.				

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 168982)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens, clone RP11-114M4
Unpublished
2 (bases 1 to 168982)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,

Anderson, S., Barak, H., Barak, N., Beckwith, K., Beda, F., Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G., Castle, A., Choepel, Y., Collangelo, M., Collins, S., Collymore, A., Cooke, P., Darcangelo, K., Dewar, K., Domino, M., Doyle, M., Fennestor, J., Ferreira, P., Fitzhugh, W., Forrest, C., Gage, D., Galan, J., Gardyna, S., Grant, G., Hagos, B., Healdorf, A., Horton, L., Howland, J. C., Johnson, R., Jones, C., Kann, L., Kartas, A., Klein, J., Landers, T., Leloczky, D., Levine, R., Lieu, C., Liu, G., Locke, K., MacDonald, P., Margulis, N., McMan, P., McGuck, A., McKenna, K., Mopheters, R., Meldrum, J., Menneus, L., Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., Olivar, T. M., Peterson, K., Pierre, N., Pisani, C., Pollard, V., Raymond, C., Riley, R., Rochman, D., Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, I., Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wymann, D., Ye, W. J., Zimmer, A. and Zody, M.

TITLE	Direct Submission
JOURNAL	Submitted (08-JAN-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 0214, USA
COMMENT	On Feb 28, 2000 this sequence version replaced gi:6751804.

Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu
 Project Information
 Center project name: L1948
 Center clone name: 114_M_4
 Summary Statistics
 Sequencing vector: M13; M77815; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 167391 bases at least Q40
 Consensus quality: 168285 bases at least Q30
 Consensus quality: 168305 bases at least Q20
 Insert size: 168000; agarose-fp
 Insert size: 168582; sum-of-contigs
 Quality coverage: 6.0 in Q20 bases; agarose-fp
 Quality coverage: 6.0 in Q20 bases; sum-of-contigs

Db 30174 TGGATATGCAATTGAGAAAAAAGTAATTTAGCTCTCCACATACCATATCTCGAAT 30233
 Qy 243 ctactactagaagaagtcacacatgltgtgaactcatctagactcagatccatcgttag 302
 Db 30234 TAATTCAGATGCAATTTTACACTAATATGGAAGATAGACACAGAGCTCTATAGAGA 30293
 Qy 303 tcaatttagcagatgaatcgtgaacagcgtccacagaatttttagcagaattgttaca 362
 Db 30294 AAGACATGGAGATTAATTTATAGTAGTCTTAAGCTAGATTCAGAAAGCAGTAACATA 30353

RESULT 14
 AC087232/c 49623 bp DNA linear INV 18-OCT-2001
 LOCUS Caenorhabditis elegans cosmid Y92H12BR, complete sequence.
 DEFINITION AC087232
 AC087232
 AC087232.2 GI:12863250
 KEYWORDS HTG.
 SOURCE Caenorhabditis elegans.
 ORGANISM Caenorhabditis elegans.
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
 Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
 1 (bases 1 to 49623)
 The C. elegans Sequencing Consortium.
 Genome sequence of the nematode C. elegans: a platform for
 investigating biology. The C. elegans Sequencing Consortium
 Science 282 (5396), 2012-2018 (1998)
 99069613
 2 (bases 1 to 49623)
 Tin-Wollam, A., Wollmann, P. and Courtney, L.
 The sequence of C. elegans cosmid Y92H12BR
 Unpublished
 3 (bases 1 to 49623)
 Waterston, R.
 Direct Submission
 Unpublished
 4 (bases 1 to 49623)
 Waterston, R.
 Direct Submission
 Submitted (19-DEC-2000) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 5 (bases 1 to 49623)
 Waterston, R.
 Direct Submission
 Submitted (16-FEB-2001) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 6 (bases 1 to 49623)
 Waterston, R.
 Direct Submission
 Submitted (23-JUN-2001) Department of Genetics, Washington
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 7 (bases 1 to 49623)
 Waterston, R.
 Direct Submission
 Submitted (18-OCT-2001) Department of Genetics, Washington
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 On Feb 16, 2001 this sequence version replaced gi:11890811.
 Submitted by:
 Genome Sequencing Center
 Department of Genetics, Washington University
 St. Louis, MO 63110, USA, and
 Sanger Centre, Hinxton Hall
 Cambridge CB10 1HQ, England
 email: rtw@emabode.wustl.edu and jess@sanger.ac.uk

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

NOTICE: This sequence may not be the entire insert of this clone.
 It may be shorter because we only sequence overlapping sections
 once, or longer because we provide a small overlap between
 neighboring submissions.

This sequence was finished as follows unless otherwise noted: all

FEATURES

source

regions were double stranded, sequenced with an alternate chemistry
 or covered by high quality data (i.e., pired quality >= 30); an
 attempt was made to resolve all sequencing problems, such as
 compressions and repeats; all regions were covered by sequence from
 more than one ml3 subclone.

NOTES:

Coding sequences below are predicted from computer analysis, using
 the program Genefinder (P. Green and L. Hillier, ms in preparation).

Location/Qualifiers

1..49623

/organism="Caenorhabditis elegans"

/strain="Bristol N2"

/db_xref="taxon:6239"

/chromosome="I"

/clone="Y92H12BR"

2892..11265

/gene="Y92H12BR.7"

join(2892..3103,4082..4220,4262..4484,5839..6030,
7491..7670,10415..11265)

/gene="Y92H12BR.7"

/codon_start=1

/product="Hypothetical protein Y92H12BR.7"

/protein_id="AAK09079.1"

/db_xref="GI:12863254"

/translation="MSPNGMQPMPHNOASRPDSMPVANNNGGPIROPPOAORS
NCMLVLANPYFSKRVGAQSHYTYNERRRMVFSPOPYVQKPLISQAPPPNPSEF
VQTLMSKRIPEPPPTPMFLDSVRSKCEMSNQKLEKERRKQMWQFQSFQF
EPKLSFDGSSSRREVCGAFELFVSPDRHDXIDFLTSTHTKTKMTSEDPQST
KNVELQVATAGDIYEIKCISTDKDFLVAIWSKEIRAEFSYSELRLQKCYOFQRG
IPNAEPLMEFECVVRCKRYKRWDEIHHCNCNVOMOKYIKDNGCGNLINGKILFLF
SARLPDLITLPOFSPSPGNRMVINCVELINPERHNFVADFCNKHIVYVCLVN
SETDICYREKLYKMLNPSPYKLYPPVYDQGRFVYNTLYWELTYTEDIQDNG
QPSALMATAGTSYRIGLDPNNQCTMCMLYPIGKKLLEPAVPEVASVKNVDSMTTI
LREGGHDSEVADTLAYLIDREVERSLTAQELAKRDLDELKAVDHNKAVAREKNA
SMTSIFRNLDNVEVTGFENKRDALFEERIKLNK"

14267..21355

/gene="Y92H12BR.6"

join(14267..14887,20972..21355)

/gene="Y92H12BR.6"

/note="coded for by the following C. elegans cDNAs:
yk258f5.5, yk372d4.5, yk258f5.3"

/codon_start=1

/product="Hypothetical protein Y92H12BR.6"

/protein_id="AAK09076.2"

/db_xref="GI:16259269"

/translation="MNEFISWCKGRGYOPDGLIYCPCPPNGCGNGIYATGFRGKAF
ITLPETDMINALVVDLPYVRKKLAKIGGKLEKPEILTMFEAFEDTSHSAMPYLV
LPKENTPAFGIDVDVNTLPISIKRYMDOKKEISEKVKSVTKNPIFLRSGV
FLQSSKFRKRIAKNFNSPIFMFTORAKRKRIIAKNSNFRFEMWTREHVELEE
VGLPSHLYATDEGSPWSLRTNIRVMLLHQMANKRMKIIYSHDPIDEDEKIDEX
LEIMLEKLRNGVAERAKKMPIDVKKWMEQYKVCVVALQSEFKKLEKLEKLEEYS
"

complement(22391..24116)

/gene="Y92H12BR.2"

complement(join(22391..22562,23561..23672,23723..23810,
23876..23969,24049..24116))

/gene="Y92H12BR.2"

/note="coded for by the following C. elegans cDNAs:
yk114a10.3, yk114a10.5, yk254e10.3, yk254e10.5, yk337h7.3,
yk337h7.5, yk605b11.3, yk605b11.5"

/codon_start=1

/product="Hypothetical protein Y92H12BR.2"

/protein_id="AAK09077.1"

/db_xref="GI:12863252"

/translation="MTTRKVPKSIYVAGSSKITLNDRESAIPKPVKATVMAKQPKR
SAQPKRFSVEEVEYVPMKPRPKRPFIOQARVITQIRVSPVNMVPOREVPRI
NNEVRSNOMONNEFKKPERONFNNHNRKMFNGNNRPHQKOPFPAKKSVAELD
RELDYMRKPKHAPIT"

24914..32931

/gene="Y92H12BR.3"

join(24914..25060,25118..25283,28534..29631,31011..31315,

gene
CDS

Db 35243 ACATCCAGATAGTGATCAAGAA 35219

Search completed: May 22, 2002, 06:36:01
Job time: 7746 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 22, 2002, 04:29:50 ; Search time 463.88 Seconds
(Without alignments)
1850.602 Million cell updates/sec

Title: US-09-693-205-7_COPY_1_500
Perfect score: 500
Sequence: 1 atgattacagaagaacatc.....gcagaattgtgtaacaa 500

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

N.Geneseq_032802:.*
1: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	500	100.0	12792	22	AAH20176 Human mutated spas
2	500	100.0	12793	22	AAH20174 Human spastin nucl
3	500	100.0	12793	22	AAH20178 Human mutated spas
4	500	100.0	12793	22	AAH20179 Human mutated spas
5	500	100.0	12793	22	AAH20182 Human mutated spas
6	331.2	66.2	11493	22	AAH20175 Mouse spastin nucl
7	42.2	8.4	1251	21	AAH70119 Plasmodium falcipar
8	38.4	7.7	1632	22	AAH32525 Human secreted pro
9	38.4	7.7	1892	22	AAH02544 Human secreted pro

C	10	38.4	7.7	3644	23	ABL12784	Drosophila melanog
C	11	37.8	7.6	513445	22	AA161373	Soybean 318013 reg
C	12	37.4	7.5	79457	23	ABL26428	Drosophila melanog
C	13	37.4	7.5	11172	24	ABL34053	Human immune syste
C	14	37.2	7.4	3129	23	AAH55610	Streptococcus pneu
C	15	37.2	7.4	39003	22	AAH28534	Genomic fragment #
C	16	37.2	7.4	40862	22	ABL34073	Human immune syste
C	17	36	7.2	11410	20	AAH13127	Enterococcus faeca
C	18	36	7.2	1038602	20	AAH01425	Complete genome se
C	19	35.6	7.1	516	20	AAH99579	Nucleic acid seque
C	20	35.6	7.1	2531	19	AAH42968	Streptococcus pneu
C	21	35.6	7.1	3102	22	AAH90796	CFE 100 coding seq
C	22	35.6	7.1	3129	22	AAH91296	Streptococcus pneu
C	23	35.6	7.1	19446	19	AAH52184	Streptococcus pneu
C	24	35.4	7.1	735	22	AAH96604	Human neuroblastom
C	25	35.4	7.1	1664976	19	AAH21209	Methanococcus jann
C	26	35.2	7.0	1556	22	AAH05528	Human secreted pro
C	27	35.2	7.0	2325	22	AAH15284	Human cDNA sequenc
C	28	35.2	7.0	2895	22	AAH05496	Human secreted pro
C	29	35	7.0	965	20	AAH13459	Enterococcus faeca
C	30	35	7.0	6536	24	ABL32147	Human immune syste
C	31	35	7.0	8378	22	AAH55763	Human adult form o
C	32	35	7.0	8378	22	AAH55764	Human neonatal for
C	33	35	7.0	32185	22	AAH34542	Human DNA for a no
C	34	34.8	7.0	427	22	AAH25826	Human breast cance
C	35	34.8	7.0	505	22	AAH18992	Human breast cance
C	36	34.8	7.0	544	23	AAH52216	Staphylococcus aur
C	37	34.8	7.0	546	23	AAH54963	Staphylococcus aur
C	38	34.8	7.0	658	21	AAH28091	Human colon cancer
C	39	34.8	7.0	744	22	AAH91766	Human cDNA 5'-end
C	40	34.8	7.0	744	22	AAH93599	Human cDNA clone r
C	41	34.8	7.0	842	18	AAH75119	Staphylococcus aur
C	42	34.8	7.0	1323	22	AAH72783	Human cervical can
C	43	34.8	7.0	1635	22	AAH94210	Human full-length
C	44	34.8	7.0	1711	22	AAH75817	Human G protein 44
C	45	34.8	7.0	1724	21	AAH16694	Human secreted pro

ALIGNMENTS

RESULT	1	ALIGNMENTS
AAH20176	AAH20176 standard; DNA; 12792 BP.	
ID	AAH20176	
AC	AAH20176;	
XX	09-AUG-2001 (first entry)	
DT		
XX		
DE	Human mutated spastin nucleotide sequence SEQ ID NO:7.	
XX		
KW	Human; mouse; spastin; ARSACS; chromosome 13q11; identification;	
KW	autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;	
KW	neurodegenerative disease; reduced sensory nerve conduction; diagnosis;	
KW	reduced motor nerve velocity; hypermyelination of retinal nerve fibre;	
KW	atrophy of upper cerebellar vermis; absence of purkinje cell;	
KW	abnormal neuronal lipid storage; genetic disorder; characterisation; ds.	
XX		
OS	Homo sapiens.	
XX		
OS	Synthetic.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	77..6604
FT		/*tag= a
FT		/product= "mutated spastin"
PN	W0200129266-A2.	
XX		
PD	26-APR-2001.	
XX		
PF	20-OCT-2000; 2000WO-US29130.	
XX		
PR	20-OCT-1999; 99US-0160588.	

XX (UYMC-) UNIV MCGILL.
PA (HOPI-) HOPITAL SAINTE-JUSTINE.
XX
XX
PI Hudson TJ, Engert J, Richter A;
XX
XX WPI: 2001-308494/32.
DR P-PSDB: AAB97821.
DR
XX
XX New isolated polynucleotide, encoding spastin gene, and polypeptides,
PT useful for diagnosing autosomal recessive spastic ataxia of
PT Charlevoix-Saguenay disease by detecting two point mutations in spastin
PT gene sequence -
XX
XX
XX Claim 1; Page -: 76pp; English.

CC The present invention describes human and mouse spastin, and mutated
 CC human spastin (autosomal recessive spastic ataxia of Charlevoix-Seguenay
 CC (ARSAOS)) gene sequences (1). The spastin gene has been mapped to
 CC chromosome 13q11. (1) have neuroprotective activities and can be used in
 CC gene therapy and as a spastin polypeptide agonists. (1), their fragments
 CC or their complements can be useful for assaying the presence of a nucleic
 CC acid molecule in a sample. (1) is useful for diagnosing or aiding in the
 CC diagnosis of an early onset neurodegenerative disease in an individual.
 CC The neurodegenerative disease comprises reduced sensory nerve conduction,
 CC reduced motor nerve velocity, hypermyelination of retinal nerve fibres,
 CC atrophy of upper cerebellar vermis, absence of Purkinje cells and
 CC abnormal neuronal lipid storage. (1) can also be used to produce
 CC antisense nucleic acids, is useful as molecular weight or chromosome
 CC markers, to identify genetic disorders, as hybridisation probes or
 CC primers, as an antigen, identify and express recombinant protein for
 CC analysis, characterisation or therapeutic use, or as markers for tissues
 CC in which the corresponding protein is expressed. Diagnostic methods from
 CC the present invention can be used to identify subjects having or at risk
 CC of developing a disease or disorder associated with aberrant expression
 CC or activity of (1). The assays can be utilised to identify a subject
 CC having or at risk of developing a disorder associated with spastin
 CC protein or spastin gene expression or activity. The present sequence
 CC encodes a mutated human spastin.
 CC N.B. The present sequence is not given in the present specification but
 CC is derived from the human spastin nucleotide sequence (AAH20174) as
 CC stated on page 14.

Sequence 12792 BP; 4163 A; 2256 C; 2487 G; 3886 T; 0 other;

Query Match	100.08;	Score 500;	DB 22;	Length 12792;
-------------	---------	------------	--------	---------------

Matches	500;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
---------	------	--------------	----	------------	----	--------	----	------	----

QY	1	atgatttaacgaagaacacatgtactacagctgcgaacgtctctaaatccaaagaattgacg	60
Db	1	atgatttaacgaagaacacatgtactacagctgcgaacgtctctaaatccaaagaattgacg	60
QY	61	tcttatcaaggaagtaatgaatacatctctggtcgtgcagagaattgattgttccaaigtta	120
Db	61	tcttatcaaggaagtaatgaatacatctctggtcgtgcagagaattgattgttccaaigtta	120
QY	121	tcgatttataaagacgaatacgcacctcgtttcctgaatggtcttaagaatgtttggaaaaa	180
Db	121	tcgatttataaagacgaatacgcacctcgtttcctgaatggtcttaagaatgtttggaaaaa	180
QY	181	tctttataatcatctttccagaagatttgaaatttgaatgaagatgcacctatccccag	240
Db	181	tctttataatcatctttccagaagatttgaaatttgaatgaagatgcacctatccccag	240
QY	241	aactatactagaagaagctcagacatgtgtggaacctattagactacgcaggttccalcgtt	300
Db	241	aactatactagaagaagctcagacatgtgtggaacctattagactacgcaggttccalcgtt	300
QY	301	agtcattttagacgaatcgtgaagcacagcttccagaattttttagacagacattgaca	360
Db	301	agtcattttagacgaatcgtgaagcacagcttccagaattttttagacagacattgaca	360

Accession	Sequence	Length
Qy	aaacttggaggggtttgtccttaaaaaattagatgactctatacaatccggttttaa	420
Db	aaacttggaggggtttgtccttaaaaaattagatgactctatacaatccggttttaa	420
Qy	aaatatctatccatccattacaagtgctgtttgcagataatggagaagatgccatt	480
Db	aaatatctatccatccattacaagtgctgtttgcagataatggagaagatgccatt	480
Qy	gcagaatgtgttaatcaaa 500	500
Db	gcagaatgtgttaatcaaa 500	500

RESULT 2

ID AAH20174 standard; DNA; 12793 BP.

ID	AAH
----	-----

AC AAH20174;

DT 09-AUG-2001 (first entry)

DE Human spastin nucleotide sequence SEQ ID NO:1

KW Human; mouse; spastin; ARSACS; chromosome 13q11; identification;
 KW autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;
 KW neurodegenerative disease; reduced sensory nerve conduction; diagnosis;
 KW reduced motor nerve velocity; hypermyelination of retinal nerve fibre;
 KW atrophy of upper cerebellar vermis; absence of Purkinje cell;
 KW abnormal neuronal lipid storage; genetic disorder; characterisation; ds
 XX
 OS Homo sapiens

Homo sapiens.

FH	Key	Location/Qualifiers
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7
8	8	8
9	9	9
10	10	10
11	11	11
12	12	12
13	13	13
14	14	14
15	15	15
16	16	16
17	17	17
18	18	18
19	19	19
20	20	20
21	21	21
22	22	22
23	23	23
24	24	24
25	25	25
26	26	26
27	27	27
28	28	28
29	29	29
30	30	30
31	31	31
32	32	32
33	33	33
34	34	34
35	35	35
36	36	36
37	37	37
38	38	38
39	39	39
40	40	40
41	41	41
42	42	42
43	43	43
44	44	44
45	45	45
46	46	46
47	47	47
48	48	48
49	49	49
50	50	50
51	51	51
52	52	52
53	53	53
54	54	54
55	55	55
56	56	56
57	57	57
58	58	58
59	59	59
60	60	60
61	61	61
62	62	62
63	63	63
64	64	64
65	65	65
66	66	66
67	67	67
68	68	68
69	69	69
70	70	70
71	71	71
72	72	72
73	73	73
74	74	74
75	75	75
76	76	76
77	77	77
78	78	78
79	79	79
80	80	80
81	81	81
82	82	82
83	83	83
84	84	84
85	85	85
86	86	86
87	87	87
88	88	88
89	89	89
90	90	90
91	91	91
92	92	92
93	93	93
94	94	94
95	95	95
96	96	96
97	97	97
98	98	98
99	99	99
100	100	100

FT / *tag= a

XX

WO200129266-A2

PD 26-APR-2001.

PF 20-OCT-2000; 2000WO-US29130.

PR 20-OCT-1999; 99US-0160588.

PA (UYMC-) UNIV MCGILL.

PA (HOPI-) HOPITAL SAINTE-JUSTINE.

PI Hudson TJ, Engert J, Richter A;

DR WPI; 2001-308494/32.

DR P-PSDB; AAB97819.

PT New isolated polynucleotide, encoding spastin gene, and polypeptides.

PT Charlevoix-Saguenay disease by detecting two point mutations in spastin.

PT gene sequ

PS Claim 1; Fig 9; 76pp; English.

CC The present invention describes human and mouse spastin and mutated
CC human spastin (autosomal recessive spastic ataxia of Charlevoix-Seguenay
CC (ASACS)) gene sequences (1). The spastin gene has been mapped to
CC chromosome 13q11. (1) have neuroprotective activities and can be used in
CC gene therapy and as a spastin polypeptide agonists. (1), their fragments
CC or their complements can be useful for assaying the presence of a nucleic
CC acid molecule in a sample. (1) is useful for diagnosing or aiding in the
CC diagnosis of an early onset neurodegenerative disease in an individual.
CC The neurodegenerative disease comprises reduced sensory nerve conduction,
CC reduced motor nerve velocity, hypermyelination of retinal nerve fibres,
CC atrophy of upper cerebellar vermis, absence of Purkinje cells and
CC abnormal neuronal lipid storage. (1) can also be used to produce

KM Human: mouse; spastin; ARSACS; chromosome 13q11; identification;
 KM autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;
 KM neurodegenerative disease; reduced sensory nerve conduction; diagnosis;
 KM reduced motor nerve velocity; hypermyelination of retinal nerve fibre;
 KM atrophy of upper cerebellar vermis; absence of Purkinje cell;
 KM abnormal neuronal lipid storage; genetic disorder; characterisation; ds.
 OS
 OS Homo sapiens.
 OS Synthetic.
 FH Key Location/Qualifiers
 FT CDS 77..11566
 FT /*tag= a
 FT /product= "mutated spastin"
 FT
 FT
 PN W0200129266-A2.
 PD 26-APR-2001.
 XX
 XX 20-OCT-2000; 2000MO-US29130.
 PF
 XX 20-OCT-1999; 99US-0160588.
 PR
 XX (UYMC-) UNIV MCGILL.
 PA (HOP-) HOPITAL SAINTE-JUSTINE.
 XX
 PI Hudson TJ, Richter A;
 XX
 DR WPI: 2001-308494/32.
 DR P-PSDB: AAB97823.
 XX
 PT New isolated polynucleotide, encoding spastin gene, and polypeptides,
 PT useful for diagnosing autosomal recessive spastic ataxia of
 PT Charlevoix-Saguenay disease by detecting two point mutations in spastin
 PT gene sequence -
 PS
 PS Claim 1: Page -: 76pp; English.
 XX
 CC The present invention describes human and mouse spastin, and mutated
 CC human spastin (autosomal recessive spastic ataxia of Charlevoix-Saguenay
 CC (ARSACS)) gene sequences (I). The spastin gene has been mapped to
 CC chromosome 13q11. (I) have neuroprotective activities and can be used in
 CC gene therapy and as a spastin polypeptide agonists. (II), their fragments
 CC or their complements can be useful for assaying the presence of a nucleic
 CC acid molecule in a sample. (II) is useful for diagnosing or aiding in the
 CC diagnosis of an early onset neurodegenerative disease in an individual.
 CC The neurodegenerative disease comprises reduced sensory nerve conduction,
 CC reduced motor nerve velocity, hypermyelination of retinal nerve fibres,
 CC atrophy of upper cerebellar vermis, absence of Purkinje cells and
 CC abnormal neuronal lipid storage. (I) can also be used to produce
 CC antisense nucleic acids, is useful as molecular weight or chromosome
 CC markers, to identify genetic disorders, as hybridisation probes or
 CC primers, as an antigen, identify and express recombinant protein for
 CC analysis, characterisation or therapeutic use, or as markers for tissues
 CC in which the corresponding protein is expressed. Diagnostic methods from
 CC the present invention can be used to identify subjects having or at risk
 CC of developing a disease or disorder associated with aberrant expression
 CC or activity of (I). The assays can be utilised to identify a subject
 CC having or at risk of developing a disorder associated with Spastin
 CC protein or spastin gene expression or activity. The present sequence
 CC encodes a mutated human spastin.
 CC N.B. The present sequence is not given in the present specification but
 CC is derived from the human spastin nucleotide sequence (AAH20174) as
 CC stated on page 14.
 XX
 XX Sequence 12793 BP: 4163 A; 2257 C; 2487 G; 3886 T; 0 other;
 SQ

Query Match 100.0%; Score 500; DB 22; Length 12793;
 Best Local Similarity 100.0%; Pred. No. 8.6e-127;
 Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 atgattacagaagacatgactcagctcagcttcaatccagaagattgcacg 60

Db
 1 atgattacagaagacatgactcagctcagcttcaatccagaagattgcacg 60
 QY
 61 tctatacaagaagtaatgatacatctcgtgcgcgagaagattgattcaatgta 120
 Db
 61 tctatacaagaagtaatgatacatctcgtgcgcgagaagattgattcaatgta 120
 QY
 121 tccatttgatgaagaacagaaatcccatctgttcatcgttgaagatggtttgaa 180
 Db
 121 tccatttgatgaagaacagaaatcccatctgttcatcgttgaagatggtttgaa 180
 QY
 181 tcttatacatcttccagaggaattgacttatttgatagatgacccatccacag 240
 Db
 181 tcttatacatcttccagaggaattgacttatttgatagatgacccatccacag 240
 QY
 241 aactatacagaagaagtcagacatgtgtgaaactaatgactcagaatccatcgt 300
 Db
 241 aactatacagaagaagtcagacatgtgtgaaactaatgactcagaatccatcgt 300
 QY
 301 agtcatttgacgatgaatctgagcacagcttccagaatttttagagacattgta 360
 Db
 301 agtcatttgacgatgaatctgagcacagcttccagaatttttagagacattgta 360
 QY
 361 aaaacttgagaggttctccttaaaaattagatgacatctatacaacatccgtat 420
 Db
 361 aaaacttgagaggttctccttaaaaattagatgacatctatacaacatccgtat 420
 QY
 421 aaatatattcattccatccatccaagtgctgtttgcagataatgaggaagatgca 480
 Db
 421 aaatatattcattccatccatccaagtgctgtttgcagataatgaggaagatgca 480
 QY
 481 gcagaaatgtgtatcaaa 500
 Db
 481 gcagaaatgtgtatcaaa 500
 RESULT 6
 AAH20175
 ID AAH20175 standard; DNA: 11493 BP.
 AC AAH20175;
 XX
 XX 09-AUG-2001 (first entry)
 DE Mouse spastin nucleotide sequence SEQ ID NO:3.
 XX
 XX Human: mouse; spastin; ARSACS; chromosome 13q11; identification;
 KM autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;
 KM neurodegenerative disease; reduced sensory nerve conduction; diagnosis;
 KM reduced motor nerve velocity; hypermyelination of retinal nerve fibre;
 KM atrophy of upper cerebellar vermis; absence of Purkinje cell;
 KM abnormal neuronal lipid storage; genetic disorder; characterisation; ds.
 OS
 OS Mus musculus.
 FH Key Location/Qualifiers
 FT CDS 1..11493
 FT /*tag= a
 FT /product= "spastin"
 FT
 PN W0200129266-A2.
 PD 26-APR-2001.
 XX
 XX 20-OCT-2000; 2000MO-US29130.
 PF
 XX 20-OCT-1999; 99US-0160588.
 PR
 XX (UYMC-) UNIV MCGILL.
 PA (HOP-) HOPITAL SAINTE-JUSTINE.
 XX
 PI Hudson TJ, Richter A;
 XX

Query Match	7.7%	Score 38.4	DB 22	Length 1632
Best Local Similarity	50.5%	Pred. No. 1		
Matches 93	Conservative 0	Mismatches 91	Indels 0	Gaps 0
Qy 26	cgagcgcagcttctaaaccagaagattgcagctctatcagaagaagtaataca	85		
Db 14172	CAGATACATGATTTAACTGATGACCTAAACCACGTTTAAAGATTATTTCTGGAAAC	1413		
Qy 86	tcttgccctgcagagaagtgtgttcaatggtatccatttgaagaaacagaatcac	145		
Db 1412	TATACACTTTTACCTTTATGACCTTTCAGAAAGTACATGATTTGTAATAATTCAGTTT	1353		
Qy 146	ccatctgttcaatgagcttaagaatggtttggaaaactttatatacatlittcagagat	205		
Db 1352	CCATCTGTTACATAGTATATATCTAGTCCATTAATCCAGTAAATCTGTTTCATGATCAC	1293		
Qy 206	ttga 209			
Db 1292	CTCA 1289			

Sequence 1632 BP; 375 A; 340 C; 387 G; 530 T; 0 other;

skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders, angiogenic disorders, kidney disorders, gastrointestinal disorders, pregnancy-related disorders, endocrine disorders, and infections. The proteins can also be used to aid wound healing and epithelial cell proliferation, to prevent skin aging due to sunburn, to maintain organs before transplantation, for supporting cell culture of primary tissues, to regenerate tissues, to identify their cognate ligands or binding partners, and in chemotaxis, and can be used as a food additive or preservative to modify storage properties. Antibodies specific for a protein of the invention can be used in alleviating symptoms associated with the disorders mentioned above, and in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked immunosorbent assay (ELISA). The present sequence represents a human secreted protein-encoding cDNA of the invention.

Human secreted protein gene #25.

Human secreted protein; diagnosis: autoimmune disease; ss; rheumatoid arthritis; hyperproliferative disorder; neoplasm; sunburn; cardiovascular disorder; cardiac arrest; cerebrovascular disorder; cerebral ischaemia; angiogenesis; nervous system disorder; skin aging; Alzheimer's disease; infection; ocular disorder; corneal infection; wound healing; epithelial cell proliferation; chemotaxis; preservative; organ transplantation; tissue regeneration; food additive.

Homosapiens.

MO200123409-A2.

05-APR-2001.

26-SEP-2000; 2000MO-US26371.

27-SEP-1999; 99US-0155804.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Ruden SM, Komatsoulis GA;

WPI, 2001-266139/27.

P-PSDB; AAU01636.

XX Nucleic acids encoding 38 human secreted polypeptides, useful for
 PT preventing, diagnosing and/or treating e.g. cancers, Parkinson's
 PT disease and diabetic retinopathy -
 XX
 PS Disclosure; Page 435; 488pp; English.
 XX
 CC AAS02511-AAS02557 represent human secreted protein coding sequences
 CC and primers of the invention. The human secreted protein sequences are
 CC used to prevent, treat or ameliorate a medical condition in e.g. humans,
 CC mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They are
 CC also used in diagnosing a pathological condition or susceptibility to a
 CC pathological condition. The antibodies to human secreted proteins can
 CC also be used in alleviating symptoms associated with the disorders and in
 CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
 CC immunosorbent assays (ELISA). Disorders which are diagnosed or treated
 CC include autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
 CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
 CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
 CC angioneuromatosis, nervous system disorders e.g. Alzheimer's disease,
 CC infections caused by bacteria, viruses and fungi and ocular disorders
 CC e.g. corneal infection. The polypeptides can also be used to aid wound
 CC healing and epithelial cell proliferation, to prevent skin aging due to
 CC sunburn, to maintain organs before transplantation, for supporting cell
 CC culture of primary tissues, to regenerate tissues and in chemotaxis. The
 CC polypeptides can also be used as a food additive or preservative to
 CC increase or decrease storage capabilities.
 CC
 XX Sequence 1892 BP; 447 A; 371 C; 478 G; 593 T; 3 other;

Query Match 7.7%; Score 38.4; DB 22; Length 1892;
 Best Local Similarity 50.5%; Pred. No. 1.1;
 Matches 93; Conservative 0; Mismatches 91; Indels 0; Gaps 0;
 QY 26 cagctcagcttcaataccagaagattgcagctcttcaagaagaatgaatca 85
 DB 1731 CAGATCAATGATTAAATCTAGACATPAAACCCAGTTTAAAGAAATTTCTGGAAC 1672
 QY 86 ttctggccctgcagagaattgcttgatccattgattgaagaacagaacac 145
 DB 1671 TATAGACTTTTACCTTTATGACCTTTCAGAGTACTTATGTAATTCAGTTT 1612
 QY 146 ccactcgttcaatgccttaagaatggttgaagaatccttataacatttcagaagat 205
 DB 1611 CCATCGTTTCAATGATATATAGTACGTCATTAATCCAGTAATCTGTTTCAATGATGAC 1552
 QY 206 ttga 209
 DB 1551 CTCA 1548

RESULT 10
 ABL12784/c
 ID ABL12784 standard; cDNA; 3644 BP.
 XX
 AC ABL12784;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 32834.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ss.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09231.
 XX

PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX
 DR WPI; 2001-656860/75.
 DR P-PSDB; ABB68681.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 PS Claim 1; SEQ ID NO 32834; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
 CC sequences (AB101840-AB116175) and the encoded proteins
 CC (ABBS7737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX Sequence 3644 BP; 1243 A; 728 C; 721 G; 952 T; 0 other;

Query Match 7.7%; Score 38.4; DB 23; Length 3644;
 Best Local Similarity 47.8%; Pred. No. 1.3;
 Matches 111; Conservative 0; Mismatches 121; Indels 0; Gaps 0;
 QY 215 ttgatgagatgccaacttaccagaactactagagaaggtcagacatgtgtgaa 274
 DB 362 TCTGGAGCCGATTTGCTGTGCTGATTTGGAATACGACGCTGCAATTCGAGGCG 303
 QY 275 ctcatctagactcagatcgcattcattttagaagatcgaatcgaagcagctc 334
 DB 302 CTTATGTGGCTTTGAATAGGCCCGAAGGAGCATGTGCTGATTAATTAAGCCGCTGACG 243
 QY 335 ccgaatctttagcagacatgtgacaaaacttggaaggttgccttaaaaattagat 394
 DB 242 CCAGGATTTGGTAATCGAATTTGGAAGAAGTGGGTTAAGATTCTTATTAATAATGCTGG 183
 QY 395 gcaatctaacacaatccgcttataaaaaataatcattcaccattaccaa 446
 DB 182 GAATGCAATTTAAACAAATTAATAATTAATAATTAATTAATTAATAACAA 131

RESULT 11
 AA161373/c
 ID AA161373 standard; DNA; 513445 BP.
 XX
 AC AA161373;
 XX
 DT 16-OCT-2001 (first entry)
 XX
 DE Soybean 318013 region A3, SEQ ID NO: 4.
 XX
 KW Soybean; antihelminthic; gene therapy; soybean cyst nematode; SCN;
 KW SCN resistance; rhg1; Rhg4; SCN resistant allele; plant breeding;
 KW 240017 region G3; 318013 region A3; 515002 region G2; ds.
 XX
 OS Glycine max.
 XX
 PN WO200151627-A2.
 XX
 PD 19-JUL-2001.
 XX
 PF 05-JAN-2001; 2001WO-US00552.
 XX

XX 07-JAN-2000; 2000US-0174880.
PR (MONS) MONSANTO CO.
XX
XX
XX Hauge BM, Wang ML, Parsons JD, Parnell LD;
XX
XX WPI: 2001-425872/45.
DR P-PSDB; AAM42216.
XX
XX New purified nucleic acid for producing a soybean plant having soybean
PT cyst nematode resistance and for use in plant breeding programs -
XX
XX
PS Claim 30; Page 596-893; 1353pp; English.
XX
XX The invention relates to nucleic acid molecules from regions of the
CC soybean genome which are associated with soybean cyst nematode (SCN)
CC resistance. The nucleic acids are used to transform plants, and can
CC produce soybean plants having an rhg1 or an Rhg4 SCN resistant allele.
CC The nucleic acids can be used for investigating rhg1 or Rhg4 haplotypes
CC of soybean plants and for introgressing SCN resistance or partial SCN
CC resistance into soybean plants. They can also be used in plant breeding
CC programmes. The invention also relates to proteins encoded by such
CC nucleic acid molecules, as well as antibodies capable of recognising
CC these proteins. The present sequence is a nucleic acid molecule
XX provided in the specification.
XX
SQ Sequence 513445 BP; 173367 A; 85402 C; 83912 G; 170492 T; 272 other;

Query Match 7.6%; Score 37.8; DB 22; Length 513445;
Best Local Similarity 48.4%; Pred. No. 9.8; Mismatches 112; Indels 0; Gaps 0;
Matches 10; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

QY 178 aaacttataatactttcagagagattgacttatttgatgagatgcacattccc 237
DB 477367 AATATATGTAATAATATATAGATTAAAGAAAATAATATATGCTAATGATATAA 477308
QY 238 cagagacatactagagagagtcgacatctgtggaactctttagcagatccatc 297
DB 477307 ATATATATTTATACACTATATATATTTATTTATATAGTTATTTATATTATTATTCATC 477248
QY 298 gttagcatctttagacatgacatcgaagcacagcttcagaattttagcagacatctgt 357
DB 477247 ATTAATCATATATATATATGATCAATTTACTGAACTTTATTAATAAAACTTATTAATAAAGT 477188
QY 358 acaaaacttgagaggttgcctctaaataatagat 394
DB 477187 ACATTATGTTGATTTTATTTATGATCGACATTGTAAT 477151

RESULT 12
ABL26428/c
ID ABL26428 standard; DNA: 7957 BP.
XX
XX ABL26428;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 30757.
DE
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
XX Drosophila melanogaster.
OS
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
XX
XX 23-MAR-2000; 2000US-191637P.
XX

PR 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI: 2001-656860/75.
DR
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX
PS Claim 1; SEQ ID NO 30757; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABR57737-ABR72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX
SQ Sequence 7957 BP; 2081 A; 1836 C; 1817 G; 2223 T; 0 other;

Query Match 7.5%; Score 37.4; DB 23; Length 7957;
Best Local Similarity 53.0%; Pred. No. 3.2; Mismatches 71; Indels 0; Gaps 0;
Matches 80; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 99 gagatgatgttgcattgcatcattgagaaacagaataccacatgttcat 158
DB 7286 GATAATCAATTTTGTGGAATCTATTTAATCAAGATTAAATGAATGAAT 7227
QY 159 ggcctaaagatggttggaataatccttatatacatcttcagagatctgacttattt 218
DB 7226 TGATTAGTTGCATATGTAATAATTTTACAGCAATTTTATTAATTAATGTTACTTCG 7167
QY 219 atgagatgcacattccccaactatc 249
DB 7166 AFAATTACCACTTCACAAAGTTATATAT 7136

RESULT 13
ABL34053/c
ID ABL34053 standard; DNA: 11172 BP.
XX
XX ABL34053;
XX
XX 26-MAR-2002 (first entry)
XX
XX Human immune system associated gene SEQ ID NO: 2026.
DE
XX
XX Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytosolic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antiporiatic;
KW antinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; Rheumatoid arthritis; psoriasis; bowel disease;
KW gene; ds.
XX
XX Homo sapiens.
OS
XX
XX WO200200928-A2.
XX
XX 03-JAN-2002.
XX
XX 02-JUL-2001; 2001WO-EP07537.
XX


```

XX
PR 18-JUN-1999; 99US-0140121.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Lagace RE, Patterson C, Berg KL;
XX
DR WPI; 2001-041427/05.
XX
PT Genomic library for identifying diagnostic and therapeutic
PT compositions, and for identifying virulence factors, regulatory
PT elements and drug targets, comprises Moraxella catarrhalis nucleic
PT acids -
XX
PS Claim 1; Page 171-180; 545pp; English.
XX
CC The present invention relates to a Moraxella catarrhalis genomic library
CC comprising of a combination of 41 nucleic acid molecules (see
CC AAF28514-AAF28554). The library has a number of uses described in the
CC specification e.g. is useful for identifying diagnostic and therapeutic
CC compositions. M. catarrhalis (Branhamella catarrhalis) is a large
CC aerobic, gram-negative diplococcus, normally found among the bacterial
CC flora of human upper airways. M. catarrhalis is known to cause acute,
CC localised infections such as otitis media, sinusitis and bronchopulmonary
CC infection and life-threatening, systemic diseases including endocarditis
CC and meningitis.
XX
SQ Sequence 39003 BP; 11568 A; 8751 C; 7476 G; 11208 T; 0 other;

```

```

Query Match 7.48; Score 37.2; DB 22; Length 39003;
Best Local Similarity 52.66; Pred. No. 6.1;
Matches 81; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

```

```

QY 291 ttccatcgtagtcattttagacgaatgaatcgaagcacagctccagaatttttagcag 350
   || || || || || || || || || || || || || || || || || || || ||
Db 22598 ttatctgcttgatcacaactcgtggttttaagaacacctgcacagtttaagaag 22657
QY 351 acattgtacaaaacttgcaggggttgccttaaaaaatagatgcattatcacacatc 410
   || || || || || || || || || || || || || || || || || || || ||
Db 22658 aagatggttaaatccttaataaaaaataatccttaataaaaaataatccttaataaaaaata 22717
QY 411 cgcttatataaaatatatcattcaccattacc 444
   || || || || || || || || || || || || || || || || || || || ||
Db 22718 tccctaataaaaaataatccttaataaaaaataatc 22751

```

Search completed: May 22, 2002, 06:43:32
Job time: 8022 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 22, 2002, 04:28:15 ; Search time 108.82 Seconds
(without alignments)
1128.623 Million cell updates/sec

Title: US-09-693-205-7_COPY_1_500
Perfect score: 500
Sequence: 1 atgattcagaagaacatc.....gcagaattgttaataaa 500

Scoring table:
IDENTITY_NUC
Gap 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/p10data/1/ina/5A_COMB.seq:*
2: /cgn2_6/p10data/1/ina/5B_COMB.seq:*
3: /cgn2_6/p10data/1/ina/6A_COMB.seq:*
4: /cgn2_6/p10data/1/ina/6B_COMB.seq:*
5: /cgn2_6/p10data/1/ina/PCRTUS_COMB.seq:*
6: /cgn2_6/p10data/1/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	35.6	7.1	3129	4 US-09-387-695-1	Sequence 1, Appl
C 2	34.8	7.0	658	4 US-09-328-111-775	Sequence 75, App
C 3	34.2	6.8	3456	1 US-08-190-687B-24	Sequence 24, Appl
C 4	34.2	6.8	4307	1 US-08-190-687B-7	Sequence 7, Appl
C 5	33.2	6.6	6479	2 US-08-649-046-3	Sequence 3, Appl
C 6	32.8	6.6	853	2 US-08-179-557-19	Sequence 19, Appl
C 7	32.4	6.5	3000	1 US-08-680-395-1	Sequence 1, Appl
C 8	32	6.4	149	1 US-08-629-339-4	Sequence 4, Appl
C 9	32	6.4	149	1 US-08-759-873-4	Sequence 4, Appl
C 10	32	6.4	180	3 US-09-004-113-13	Sequence 13, Appl
C 11	32	6.4	11832	2 US-08-416-603-1	Sequence 1, Appl
C 12	31.8	6.4	1419	1 US-08-242-098-39	Sequence 39, Appl
C 13	31.6	6.3	2016	4 US-09-132-118-1	Sequence 1, Appl
C 14	31.6	6.3	2106	4 US-08-923-511-1	Sequence 1, Appl
C 15	31.6	6.3	2137	1 US-08-444-005-16	Sequence 16, Appl
C 16	31.6	6.3	2617	3 US-09-161-443-1	Sequence 1, Appl
C 17	31.6	6.3	9687	4 US-09-133-944-2	Sequence 2, Appl
C 18	31.4	6.3	5014	4 US-09-381-862-3	Sequence 3, Appl
C 19	31	6.2	2848	4 US-08-936-165A-197	Sequence 197, App
C 20	31	6.2	2897	2 US-08-927-394-1	Sequence 1, Appl
C 21	31	6.2	3923	4 US-09-176-320-7	Sequence 7, Appl
C 22	31	6.2	6559	4 US-09-234-186-1	Sequence 1, Appl
C 23	31	6.2	6560	5 PCT-US93-05651-1	Sequence 1, Appl
C 24	30.8	6.2	854	3 US-08-867-381A-51	Sequence 51, Appl
C 25	30.8	6.2	854	4 US-09-521-144-51	Sequence 51, Appl
C 26	30.8	6.2	1002	4 US-08-960-780-43	Sequence 43, Appl
C 27	30.8	6.2	1002	4 US-09-073-898-43	Sequence 43, Appl

28	30.8	6.2	1500	1 US-08-117-083-67	Sequence 67, Appl
29	30.8	6.2	2239	4 US-09-196-390-1	Sequence 1, Appl
30	30.8	6.2	3000	1 US-08-184-252A-1	Sequence 1, Appl
31	30.8	6.2	3000	5 PCT-US95-00601-1	Sequence 1, Appl
32	30.6	6.1	9636	1 US-08-323-170B-1	Sequence 1, Appl
33	30.6	6.1	9636	4 US-08-954-441-1	Sequence 1, Appl
34	30.4	6.1	1491	2 US-08-218-265-9	Sequence 9, Appl
35	30.2	6.0	2750	3 US-08-617-860B-33	Sequence 33, Appl
36	30	6.0	4619	2 US-08-874-186-38	Sequence 38, Appl
37	29.8	6.0	13158	2 US-08-687-080-105	Sequence 105, App
C 38	29.6	5.9	4447	2 US-08-304-309-3	Sequence 3, Appl
C 39	29.6	5.9	40352	3 US-08-991-942-3	Sequence 3, Appl
C 40	29.6	5.9	4447	3 US-08-846-111D-15	Sequence 15, Appl
C 41	29.6	5.9	56516	2 US-08-996-306-1	Sequence 1, Appl
C 42	29.6	5.9	56516	4 US-09-338-907-1	Sequence 1, Appl
C 43	29.6	5.9	56516	4 US-09-218-207-1	Sequence 1, Appl
C 44	29.6	5.9	56520	4 US-09-338-907-179	Sequence 179, App
C 45	29.6	5.9	56520	4 US-09-218-207-179	Sequence 179, App

ALIGNMENTS

```
RESULT 1
US-09-387-695-1/c
: Sequence 1, Application US/09387695
: Patent No. 6280990
: GENERAL INFORMATION:
: APPLICANT: May, Earl
: APPLICANT: Van Horn, Stephanie
: APPLICANT: Warren, Patrick V.
: APPLICANT: Warren, Richard L.
: TITLE OF INVENTION: dnae
: FILE REFERENCE: GM10237
: CURRENT APPLICATION NUMBER: US/09/387,695
: CURRENT FILING DATE: 1999-08-31
: NUMBER OF SEQ ID NOS: 2
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 1
: LENGTH: 3129
: TYPE: DNA
: ORGANISM: Streptococcus pneumoniae
: US-09-387-695-1

Query Match 7.1%; Score 35.6; DB 4; Length 3129;
Best Local Similarity 51.9%; Pred. No. 0.5;
Matches 80; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 106 gatgttcattggtatccattgattgaaacagaaatccaccatctgtttatggtctaa 165
DB 2543 GATTCCTGCGCAACATTAATGATCTCCAAACAACCTCCCAACTCTTTCACAAATTCA 2484

QY 166 gatgttggaataatcttatcatcttttcagagattgacttattgtagat 225
DB 2483 AATGATTAGCTAATTAATTAATCTTTTGACAGATTTTTCAAATGATCAGAAAGA 2424

QY 226 gccactatcccgaaactactactagaggaagt 259
DB 2423 CCAACTTTTACCAAGGTTCTACGACGAGGAGT 2390

RESULT 2
US-09-328-111-775
: Sequence 775, Application US/09328111
: Patent No. 6262333
: GENERAL INFORMATION:
: APPLICANT: Endege, Willson O.
: APPLICANT: Steinmann, Kathleen E.
: APPLICANT: Astle, Jon H.
: APPLICANT: Burgess, Christopher C.
: APPLICANT: Bushnell, Steven E.
: APPLICANT: Carroll III, Eddie
```

APPLICANT: Catino, Theodore J.
APPLICANT: Dertl, Adnan
APPLICANT: Ford, Donna M.
APPLICANT: Lewis, Marcia E.
APPLICANT: Monahan, John E.
APPLICANT: Schlegel, Robert
TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
FILE REFERENCE: CCD-257 (US)
CURRENT APPLICATION NUMBER: US/09/328,111
CURRENT FILING DATE: 1999-06-08
EARLIER APPLICATION NUMBER: US 60/088,801
EARLIER FILING DATE: 1998-06-10
NUMBER OF SEQ ID NOS: 850
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 775
LENGTH: 658
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc.feature
LOCATION: (1)..(658)
OTHER INFORMATION: n = A,T,C or G
US-09-328-111-775

Query Match 7.0%: Score 34.8; DB 4; Length 658;
Best Local Similarity 47.0%: Pred. No. 0.48;
Matches 108; Conservative 0; Mismatches 122; Indels 0; Gaps 0;

QY 193 ttttcagaagatttgacttatttgatgagccactatccagaactatactaga 252
DB 139 tttttatccatattagtcctgtgagcattcttcaactaaccgctttaga 198
QY 253 ggaaggtcagactggtggaactcttagaccagagatccctgtagctttaga 312
DB 199 tgaatgatatacgcacgcttgaaggaagtgtagatccctatcgagataagaataaca 258
QY 313 cgaatgatcgaagcagcagcttcagaaattttagcagacattgtacaacttgagg 372
DB 259 tgaagagctcagcttaagagatgaggaatgattgggcccataatagaataa 318
QY 373 gtttgcttaaaaaaattagatgcatctacacacatccgcttataaaa 422
DB 319 ggtggtctgtagagagagagataaaaaactaaacctgatatgataataa 368

RESULT 3
US-08-190-687B-24
Sequence 24, Application US/08190687B
Patent No. 5760203

GENERAL INFORMATION:

APPLICANT: Wong, Gail L.
APPLICANT: Martin, George
APPLICANT: McCormick, Francis P.
APPLICANT: Rubinfeld, Bonnie
APPLICANT: O'Rourke, Edward C.
APPLICANT: Clark, Robin
TITLE OF INVENTION: GAP Gene Sequences
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/190,687B
FILING DATE: 02-FEB-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/774,644
FILING DATE: 11-OCT-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/260,807
FILING DATE: 21-OCT-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/230,761
FILING DATE: 10-AUG-1988
ATTORNEY/AGENT INFORMATION:
NAME: Gass, David A.
REGISTRATION NUMBER: 38,153
REFERENCE/DOCKET NUMBER: 27527/31898
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 3456 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 100..2709
US-08-190-687B-24

Query Match 6.8%: Score 34.2; DB 1; Length 3456;
Best Local Similarity 55.5%: Pred. No. 1.3;
Matches 66; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 343 ttttagcagacattgtacaaaacttgagggttctgccttaaaaaattagatgcatctat 402
DB 2100 ttttaacacacacttttgacacttcttcagagccttggtggagaaatttgccttcagag 2159
QY 403 acacacatcgccttataaaaaatattcattaccattaccagaatgctgtttgcaga 461
DB 2160 aatacttccacgcagacattgagatattattatgggtgtttacagaaattcgttcacgata 2218

RESULT 4
US-08-190-687B-7
Sequence 7, Application US/08190687B
Patent No. 5760203

GENERAL INFORMATION:

APPLICANT: Wong, Gail L.
APPLICANT: Martin, George
APPLICANT: McCormick, Francis P.
APPLICANT: Rubinfeld, Bonnie
APPLICANT: O'Rourke, Edward C.
APPLICANT: Clark, Robin
TITLE OF INVENTION: GAP Gene Sequences
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/08/190,687B

FILING DATE: 02-FEB-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/774,644
FILING DATE: 11-OCT-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/260,807
FILING DATE: 21-OCT-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/230,761
FILING DATE: 10-AUG-1988
ATTORNEY/AGENT INFORMATION:
NAME: Gass, David A.
REGISTRATION NUMBER: 38,153
REFERENCE/DOCKET NUMBER: 27527/31898
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ. ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 4307 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 119..3259
US-08-190-687B-7

Query Match 6.8%; Score 34.2; DB 1; Length 4307;
Best Local Similarity 55.5%; Pred. No. 1.4;
Matches 66; Conservative 0; Mismatches 53; Indels 0; Gaps 0;
Qy 343 tttagcgacatgtgtacaaactgtgaggtgtgtccttaaaatagatgcatcat 402
Db 2650 ttttaacacacctattgaacattcttcagaccttgagagaaattatcagcttcaga 2709
Qy 403 acaacacgcgttataaaatatattcattcaccaatcaagaagtgcgtttgcaga 461
Db 2710 aattcttccacccgacattgagatattttatgsgcttttacagaaactgttcagcaca 2768
RESULT 5
US-08-649-046-3/c
Sequence 3, Application US/08649046
Patent No. 5912415
GENERAL INFORMATION:
APPLICANT: JACOBSEN, STEVEN E.
TITLE OF INVENTION: THE SPINDLY GENE, METHODS OF
TITLE OF INVENTION: IDENTIFICATION AND USE
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: MUEHLING, RAASCH, GEBHARDT & SCHWAPPACH, P.A.
STREET: 119 NORTH FOURTH STREET, SUITE 203
CITY: MINNEAPOLIS
STATE: MINNESOTA
COUNTRY: USA
ZIP: 55401
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/649,046
FILING DATE: 16-MAY-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: MCCORMACK, MYRA H.

REGISTRATION NUMBER: 36,602
REFERENCE/DOCKET NUMBER: 110.00340101
TELECOMMUNICATION INFORMATION:
TELEPHONE: 612-305-1225
TELEFAX: 612-305-1228
INFORMATION FOR SEQ. ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 6479 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-649-046-3

Query Match 6.6%; Score 33.2; DB 2; Length 6479;
Best Local Similarity 55.1%; Pred. No. 3.2;
Matches 65; Conservative 0; Mismatches 53; Indels 0; Gaps 0;
Qy 382 taaaaatagatgcatctatacaacatccgttataaaatatattcattccactt 441
Db 4476 ttaaaatttaagatttttttttgagaccttccgaaccgaattgcatcttgaaaccaa 4417
Qy 442 accaagtgcgttttcgagataatggaagaatgcatgcaagaattgttaacaa 499
Db 4416 acgtactttttgttgatctgtgtgagatcagccacacgaatctgtaattctgtaacaa 4359

RESULT 6
US-08-179-557-19/c
Sequence 19, Application US/08179557
Patent No. 5837509
GENERAL INFORMATION:
APPLICANT: ISRAELSEN, Hans
APPLICANT: BECH HANSEN, Egon
APPLICANT: MADSEN, Soeren Michael
APPLICANT: JOHANSEN, Eric
APPLICANT: NILSSON, Dan
APPLICANT: VRANG, Astrid
TITLE OF INVENTION: Recombinant Lactic Acid Bacterium
TITLE OF INVENTION: Containing an Inserted Promoter and Method of Constructing
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W.
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/179,557
FILING DATE: 07-JAN-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DK 1579/92
FILING DATE: 30-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DK 0988/93
FILING DATE: 01-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/DK94/00004
FILING DATE: 03-JAN-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/036,681
FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768


```

CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/242,098
FILING DATE: 13-MAY-1994
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/133,390
FILING DATE: 08-OCT-1993
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30307/141/PLVI
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136

INFORMATION FOR SEQ ID NO: 39:
SEQUENCE CHARACTERISTICS:
LENGTH: 1419 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

US-08-242-098-39

Query Match 6.4%; Score 31.8; DB 1; Length 1419;
Best Local Similarity 52.7%; Pred.No. 4.6;
Matches 69; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 315 atgaatctgaagcacagcttcacgaatttttagcacagatgtgtacaaaacttggaaggt 374
      ||| ||||| ||| ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 707 AAGATTGTGAAGCGTCACCTTGAAAAATTGTGTAGCTGACATGCACGAACAAAATCTTCCTT 766

QY 375 ttgtccttaaaaatagatgcattctatacaacaatcgcttatataaaaatatatcatt 434
      ||| ||||| ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 767 TTGACCTTGAAGAGGCTGATGAAAAAGCACAACTTATAGTTAATTCAGCGCATG 826

QY 435 caccattacca 445
      | | |||
Db 827 GTCCGATGCCA 837

RESULT 13
US-09-132-118-1
Sequence 1, Application US/09132118
Patent No. 621137
GENERAL INFORMATION:
APPLICANT: BAICHWAL, VIJAY R
APPLICANT: HUANG, JIANING
APPLICANT: HSU, HAILING
APPLICANT: GOEDEL, DAVID V
TITLE OF INVENTION: RIP: NOVEL HUMAN PROTEIN INVOLVED IN
TITLE OF INVENTION: TUMOR NECROSIS FACTOR SIGNAL TRANSDUCTION, AND SCREENING
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
STREET: 75 DENISE DRIVE
CITY: HILLSBOROUGH
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94010

COMPUTER READABLE FORM:

```

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/132,118
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: OSMAN, RICHARD A.
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: T95-006-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 343-4341
TELEFAX: (650) 343-4342
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2016 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..2013
US-09-132-118-1

Query Match 6.3%; Score 31.6; DB 4; Length 2016;
Best Local Similarity 53.2%; Pred. No. 5.9;
Matches 67; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY 344 ttgagcagactgttacaacttgaggggtgttccttaaaaaatagatgcata 403
DB 832 TTTCTCGGCTTGAAGAAATTTAGGCCCTTTTATTAGTCAATTAGAAGAGTGA 891
QY 404 caacatccgcttataaaaatatattcattccacattcaagtgctgttcagata 463
DB 892 GAAGAGGAGCTGAAGAGTTTAAAGAAAGATTCAACGAAATGCGATTGAGAGCA 951

QY 464 atgagag 469
DB 952 ATGCAG 957

RESULT 14
US-08-923-511-1
Sequence 1, Application US/08923511
Patent No. 6274376
GENERAL INFORMATION:
APPLICANT: Burnham, Martin K.
APPLICANT: Fosberry, Andrew
APPLICANT: Hodgson, John
APPLICANT: Lawlor, Elizabeth
APPLICANT: Rosenberg, Martin
APPLICANT: Ward, Judith
TITLE OF INVENTION: No. 6274376el c1p1
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dechert, Price & Rhoads
STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103-2793
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/923,511
FILING DATE:

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/011888
FILING DATE: 20-FEB-1996
APPLICATION NUMBER: PCT/US97/02547
FILING DATE: 19-FEB-1997
APPLICATION NUMBER: PCT/US97/02318
FILING DATE: 19-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: Dickinson, Q. Todd
REGISTRATION NUMBER: 28,354
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215/994-2252
TELEFAX: 215/994-2222
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2106 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-923-511-1

Query Match 6.3%; Score 31.6; DB 4; Length 2106;
Best Local Similarity 60.5%; Pred. No. 6;
Matches 52; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 283 accagagatccatcgtagtcatctttagacgatgaatctgaagcacagcttcagatt 342
DB 1557 ACCGCCGTAATCCATACATCATCATCTATTGTGATGAATGAAGCAATCCACAAT 1616
QY 343 tttagcagacattgtacaaaacttg 368
DB 1617 TTTAACAATGTATTACAACTAATG 1642

RESULT 15
US-08-444-005-16
Sequence 16, Application US/08444005
Patent No. 5674734
GENERAL INFORMATION:
APPLICANT: Ledger, Phillip
APPLICANT: Seed, Brian
APPLICANT: Stanger, Ben Z.
APPLICANT: Lee, Tae-Ho
APPLICANT: Kim, Emily
TITLE OF INVENTION: CELL DEATH PROTEIN
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street, Suite 3100
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/444,005
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,164
REFERENCE/DOCKET NUMBER: 00383/026001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 22, 2002, 04:23:20 ; Search time 3619.39 Seconds
(without alignments)
1864.535 Million cell updates/sec

Title: US-09-693-205-7_COPY_1_500
Perfect score: 500
Sequence: 1 atgattacagaagacatc.....gcagaattgttaatacaaa 500

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 674847542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

EST:*
1: em_estda:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_juv:*
15: em_gss_pln:*
16: em_gss_vtl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	396.6	79.3	746	12 BH025017	BH025017 RPCI-24-3
C 2	396.6	79.3	746	12 BH126217	BH126217 RPCI-24-3
C 3	230.6	46.1	558	12 AZ649875	AZ649875 IM0519B14
C 4	212.4	42.5	241	9 BE008891	BE008891 CM4-BN016
C 5	163.2	32.6	496	12 AZ644393	AZ644393 IM0508M18
C 6	159.2	31.8	717	12 BH110525	BH110525 RPCI-24-3
C 7	142	28.4	883	12 CNS04H0G	AL261193 Tetradon
C 8	83.2	16.6	202	9 BB561400	BB561400 BB561400
C 9	66.4	13.3	170	10 BF397419	BF397419 UI-R-B52-
C 10	57	11.4	266	10 BF326199	BF326199 MRO-AN008
C 11	44	8.8	752	12 BH571801	BH571801 BOHLS54TR
C 12	43.8	8.8	440	12 AZ066617	AZ066617 RPCI-23-4
C 13	41.4	8.3	895	12 BH131081	BH131081 ENT0P16TR
C 14	41.4	8.3	896	12 BH136596	BH136596 ENTNC65TR
C 15	41.4	8.3	940	12 AZ547902	AZ547902 ENTDP11TR
C 16	41.4	8.3	1101	12 CNS0178M	AL107728 Drosophila
C 17	41	8.2	1101	12 CNS00DDH	AL069879 Drosophila

C 18	39.8	8.0	845	12 AZ679734	AZ679734 ENTNC87TR
C 19	39.8	8.0	926	12 BH146623	BH146623 ENTPTZ1TR
C 20	39.6	7.9	653	12 AZ611614	AZ611614 IM0438J01
C 21	39.6	7.9	753	12 AZ660635	AZ660635 IM0538014
C 22	39.6	7.9	928	12 AZ528035	AZ528035 ENTTC142TR
C 23	39.6	7.9	987	12 CNS014PQ	AL104456 Drosophila
C 24	39.2	7.8	642	9 BB523933	BB523933 BB523933
C 25	39.2	7.8	1026	12 CNS0122K	AL102218 Drosophila
C 26	39	7.8	649	12 A0576178	A0576178 dbx00088P
C 27	38.8	7.8	871	12 BH133334	BH133334 ENT0J63TR
C 28	38.8	7.8	1101	12 CNS00F4N	AL070032 Drosophila
C 29	38.6	7.7	299	10 BE578331	BE578331 rK10909.Y
C 30	38.6	7.7	451	10 BM027431	BM027431 GYT000078
C 31	38.4	7.7	387	9 A1275161	A1275161 qw08b06.x
C 32	38.4	7.7	420	12 A0003378	A0003378 CPG0364A
C 33	38.4	7.7	560	10 BF510708	BF510708 UT-H-B14-
C 34	38.4	7.7	804	10 BF220200	BF220200 601296755
C 35	38.2	7.6	766	12 BH495660	BH495660 BOH0M66TR
C 36	38.2	7.6	877	12 AZ671825	AZ671825 ENTXK58TR
C 37	37.8	7.6	165	9 A1544786	A1544786 fb68903.x
C 38	37.8	7.6	541	10 B1743087	B1743087 KX38606.Y
C 39	37.4	7.5	418	9 A0020427	A0020427 A0020427
C 40	37.4	7.5	683	9 A1292912	A1292912 GH15943.5
C 41	37.4	7.5	776	10 BF294246	BF294246 001PDH07
C 42	37.4	7.5	883	12 BH256414	BH256414 K600869.D
C 43	37.4	7.5	932	12 BH155613	BH155613 ENTXS92TR
C 44	37.4	7.5	1101	12 CNS0039G	AL063921 Drosophila
C 45	37.2	7.4	346	10 BF709704	BF709704 MI-P-Ay0-

ALIGNMENTS

RESULT 1
BH025017/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

BH025017
746 bp
DNA
Mus musculus
genomic clone RPCI-24-318E15
/ DNA sequence.
GI:14788481

REFERENCE
AUTHORS

Zhao, S., Nierman, W., Malek, J., Shatsman, S., Akintet, B., Levins, M., Tsegaye, G., Geer, K., Krol, M., Shvartsbeyn, A., Gebregeorgis, E., Russell, D., de Jong, P., and Fraser, C.M.
Mouse BAC End Sequences from Library RPCI-24
Unpublished (1999)
Other-GSSS: RPCI-24-318E15.TVB
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0200
Email: szhao@tigr.org

JOURNAL
COMMENT

Clones are derived from the mouse BAC library RPCI-24. For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end page: http://www.tigr.org/tigr/bac_ends/mouse/bac_end_intro.html
Plate: 318 row: E column: 15
Seq primer: SP6
Class: BAC ends.
Location/Qualifiers

FEATURES
SOURCE

1..746
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-318E15"

```
/clone_lib="RPCI-24"
/sex="Male"
/cell_type="Spleen/Brain"
/notes="Vector: pFARBAC1; Site_1: BamHI; Site_2: BamHI;
RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
library was cloned in the pFARBAC1 cloning vector at the
BamHI sites using MboI partially digested male C57BL/6J
DNA."
BASE COUNT      229 a      142 c      137 g      238 t
ORIGIN

Query Match      79.3%; Score 396.6; DB 12; Length 746;
Best Local Similarity 87.2%; Pred. No. 6.7e-94;
Matches 435; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY      2      tgaattacaggaagaccatgctactcagctgcagctcttaaatccagaagattgcagct 61
DB      596      TGATCTACAGAGACCATGTACTCAGCTGCACCTCTTAATCCAGAGCATTTGCACGT 537
QY      62      ctatcaagaagtaatgaatacatctcgtgcctgcagagaattgattgtaattgata 121
DB      536      CTATTATTAAGAATCATGATACATTTCTGGCTGCTGCGAGAGTTGGTTGCTGAGTGTAT 477
QY      122      ccaattgataaacaagaataccaccatctgtctcattgactaagatggtttgaaaaat 181
DB      476      CCATTTAGTGAAGACAAGCATCCCATCCCTTTCATGGCTTAAGATGCTTTGGAAGAAAT 417
QY      182      cttatatacatcttttcagaagattgactcttattgtagagatgccactatccacaga 241
DB      416      CTCTAATATACATTTCTCGGAAGATTGACATTTTGTGATGAGATGCCACTTATCCCTAGA 357
QY      242      actatacaggaaggaagtcagacatgtgtgaactcatcattagacagattcatcgttca 301
DB      356      ACTCTACTGAATAGAGACCAAGAGCTGTGTGGAACATCATCAGATCCATCCATCAGTA 237
QY      302      gtcaatttagacagatgaatctgaagcagcttcacagaatttttagacagacattgtaca 361
DB      296      GTCTATTTAGATGATGAAGTCAAGCTCAGACTTCCAGAAATTTTACAGATATTGTAGCAA 237
QY      362      aaactctggagggtctgtccttaaaaaatagatgcatctatacaaatccgcttataaa 421
DB      236      AAACCTTGAGGGGATGTCTCGAAGAAAGATGATACCTCATTTCCAGATCCATCTGTAAA 177
QY      422      aaatatatcatcaccattaccagagtcgtgttttcagaataatggaagaagatgcatg 481
DB      176      AATATACATATTCCTCCCACTCCCGAGTGTCTATTTCGACATATGAGAGATACCTCTA 117
QY      482      cagaatctgttatacaaaa 500
DB      116      CAGAACTTGATATCAAAA 98

RESULT 2
BH126217/c
LOCUS      BH126217
DEFINITION      RPCI-24-318C15.TJ RPCI-24 Mus musculus genomic clone RPCI-24-318C15
VERSION      BH126217
KEYWORDS      BH126217.1 GI:14969729
SOURCE      GSS.
ORGANISM      house mouse.
            Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE      1 (bases 1 to 796)
AUTHORS      Zhao,S., Nierman,M., Malek,J., Shatsman,S., Akimov,B., Levins,M.,
            Tseay,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebreyegorgis,E.,
            Russell,D., de Jong,P. and Fraser,C.M.
            Mouse BAC End Sequences from Library RPCI-24
            Unpublished (1999)
            Other_GSSs: RPCI-24-318C15.TVB
COMMENT      Contact: Shaying Zhao
```

Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-24. For BAC
library availability, please contact Pieter de Jong
(pdejong@mail.cho.org). Clones may be purchased from BACPAC
Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end
page: http://www.tigr.org/tdb/bac-ends/mouse/bac_end_intro.html
Plate: 318 row: C column: 15
Seq primer: SP6
Class: BAC ends.

FEATURES

source

1..796
Location/Qualifiers

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-318C15"
/clone_lib="RPCI-24"
/sex="Male"
/cell_type="Spleen/Brain"
/note="Vector: pFARBAC1; Site_1: BamHI; Site_2: BamHI;
RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
library was cloned in the pFARBAC1 cloning vector at the
BamHI sites using MboI partially digested male C57BL/6J
DNA."

BASE COUNT 244 a 151 c 144 g 257 t
ORIGIN

Query Match 79.3%; Score 396.6; DB 12; Length 796;
Best Local Similarity 87.2%; Pred. No. 6.8e-94;
Matches 435; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

```
QY      2      tgaattacaggaagaccatgctactcagctgcagctcttaaatccagaagattgcagct 61
DB      619      TGATCTACAGAGACCATGTACTCAGCTGCACCTCTTAATCCAGAGCATTTGCACGT 560
QY      62      ctatcaagaagtaatgaatacatctcgtgcctgcagagaattgattgttcaatgttat 121
DB      559      CTATTATTAAGAAATCATGATACATTTCTGGCTGCTGCGAGAGTGGTGTGCTGAGTGTAT 500
QY      122      ccaattgataaacaagaataccaccatctgttctcattgctgaagatggtttgaaaaat 181
DB      499      CCATTTAGTGAAGACAAGCATCCCATCCCTTTCATGGCTTAAGATGCTTTGGAAGAAAT 440
QY      182      cttatatacatcttttcagaagattgacttattttagatgagatgccaactatccacaga 241
DB      439      CTCTATATATACATTTCTCGGAAGATTTGACTTATTGATGAGATGGCAGTATTCCTTAGA 380
QY      242      actatacaggaaggtcagacatgtgtgaactcaattagactcagagattccatctgta 301
DB      379      ACTCTACTGAATAGAGACCAAGAGCTGTGTGGAACATCATCAGATCCATCCATCAGTA 320
QY      302      gtcaatttagacagatgaatctgaagcagcttcacagaatttttagacagacattgtaca 361
DB      319      GTCTATTTAGATGATGAAGTCAAGCTCAGACTTCCAGAAATTTTACAGATATTGTAGCAA 260
QY      362      aaactctggagggtctgtccttaaaaaatagatgcatctatacaaatccgcttataaa 421
DB      259      AAACCTTGAGGGGATGTCTCGAAGAAAGATGATACCTCATTTCCAGATCCATCTGTAAA 200
QY      422      aaatatatcatcaccattaccagagtcgtgttttcagaataatggaagaagatgcatg 481
DB      199      AATATACATATTCCTCCCACTCCCGAGTGTCTATTTCGACATATGAGAGATACCTCTA 140
QY      482      cagaatctgttatacaaaa 500
DB      139      CAGAACTTGATATCAAAA 121
```

RESULT	3
LOCUS	AZ649875
DEFINITION	AZ649875 558 bp DNA linear GSS 14-DEC-2000
ACCESSION	U00519B1.R Mouse 10kb plasmid U00C1M library Mus musculus genomic
VERSION	AZ649875
KEYWORDS	clone U00C1M0519B14 R, DNA sequence.
SOURCE	AZ649875.1 GI:11783794
ORGANISM	GSS. house mouse. Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Scurionathi; Muridae; Murinae; Mus. 1 (bases 1 to 558)
AUTHORS	Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamll,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly, M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weiss,R.
TITLE	Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL	Unpublished (2000)
COMMENT	Contact: Robert B. Weiss

Rm. 308, Biometrical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddun@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0519 row: B column: 14
Seq primer: CACACAGGAACACGATACAC
Class: plasmid ends
High quality sequence stop: 558.

FEATURES	Location/Qualifiers
source	1. .558

/organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="NC01M0519P14"
 /clone_1lb="Mouse 10kb plasmid UGCM14 library"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: pMD22ny; Purified genomic DNA from M.
 musculus C57BL/6J (male) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (<http://www.jax.org/resources/documents/dnares/>). The DNA
 was hydrothermally sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repaired with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were
 ligated to the blunt ends in high molar excess. The
 adaptor DNA was purified and size-selected for a 9.5 to
 10.5 kb range using preparative agarose gel
 electrophoresis. Vector DNA was prepared from a derivative
 of pMD22 (g114732114.9b/AP129072.1), a copy-number
 indelible derivative of plasmid R1. The vector was ligated
 with adaptors complementary to the insert adaptors and
 purified. The sheared, adaptor mouse DNA was annealed to
 adaptor vector DNA, and transformed into
 chemically-competent E. coli XL10-Gold (Stratagene) cells
 and selected for ampicillin resistance."

BASE COUNT	200 a	99 c	96 g	163 t
ORIGIN				

Query Match	46.1%	Score 230.6	DB 12	Length 558
Best Local Similarity	84.1%	Pred. No. 3e-50		
Matches	297	Conservative	0	Mismatches 49; Indels 7; Gaps 3;
Qy	149	tctgtttcaatgcttaagaatggttttggaataaactttatatacaattttcagagatttg	208	
Db	1	ttccctttatagcttwaataa----tttggaagaatctcrtttatatacaatttttcagaaatttg	56	

OY	209	actctatttgatgagatgccacttatccccgaactactacagggaagcgcaagatgt	268
Db	57	actttatttgatgagatccacttattcccttgaaattatgaaatgagaccacgttggt	116
OY	269	gtggaactcatlagaactcaaggatccacgcgttagtcatctttagacgatactcgaagca	328
Db	117	gtggaactcatcagactcagatccacgcgttagtcatctttagacgatactcgaagca	176
OY	329	cagcttccaagaatttttgcagacatctgtaacaaaactctgaaggctgtgctcttaaaaa	388
Db	177	cagcttctagaattttttagcagatatttttaacaaaactctg--ggagtagtgcctgtaaaaa	234
OY	389	ttagatgatcatatacaacatccgcgttatt-aaaaaatactcatlcaacattaccag	447
Db	235	ctagatfaccctctatttcacacatcactcgtgttaaaaaaaatagatttcattcccatfaccag	294
OY	448	tgctgtttttgcagataatgtagaagaatgcatcttgagaagaatttgatatacaa	500
Db	295	tgcatttttttgatgataatgtagaagaatgcatcttcactcactcactgtagtatac	347

LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	FEATURES	COMMENT
BE008891/c								
BE008891	CM-180161-040400-132-c08	BE008891	BE008891.1	EST	human.		241 bp mRNA linear	EST 05-JUN-2000
			GI:8269124					Homo sapiens CDNA, mRNA sequence.

REFERENCE
AUTHORS

Eukaryota: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiinae; Homo.
1 (bases 1 to 241)
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Britones, M. R.,

TITLE	JOURNAL	MEDLINE	COMMENT
Goldman, G. H., Carvalho, A. F., Matsukuma, A., Bala, G. S., Simpson, D. H., Brunstein, A., deOliveira, P. S., Bucher, P., Jongeneel, C. V., O'Hare, M. J., Soares, F., Brentani, R. R., Reis, L. F., de Souza, S. J. and Simpson, A. J. G.	Proc. Natl. Acad. Sci. U.S.A.	97 (7), 3491-3496	(2000)
Shingun sequencing of the human transcriptome with ORF expressed sequence tags			
Contact: Simpson A.J.G.			

Laboratory Of Cancer Genetics
Ludwig Institute For Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the PABSP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2-CM4-BN0161-040>)
400-132-d086t3-2000-04-046t4-1)
Seq primer: puc 18 forward
High quality sequence start: 6
High quality sequence stop: 241.

FEATURES

Source

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1lb="BN0161"
/dev_stage="Adult"
/name="Organ: breast_normal; Vector: puc18; Site.1: SmaI; Site.2: SmaI; A mini-library was made by cloning products derived from ORESTATS PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

```

Query Match	42.5%	Score 212.4	DB 9	Length 241
Best Local Similarity	99.5%	Pred. No. 1.4e+45		
Matches 213:	Conservative	0	Mismatches 1	Indels 0
QY	275	ctcaatgaactgaagatccatcgctatgaatttagaagatgaaictgaagacagctt	334	
Db	241	ctcatttaagacttcagattccatcgtttactcatttttagacgatgaattgaaacacagctt	182	
QY	335	ccagaattcttagagacacatgtgacaaaacttggagagttgtgccttaaaaatttagat	394	
Db	181	ccagaaattttttagcagacacatttgtacaaaaacttgagaggttggctttaaaaatttagat	122	
QY	395	gcattcatcaaacatccgctatttaaaaatatattcaacatcccaagtctgt	454	
Db	121	gcatttataacaaacatcccttt	62	
QY	455	ttgcagataatgagaaagatgccattgagaagt	488	
Db	61	ttgcagataatgagaaagatgccattgagaagt	28	

RESULT	5
AZ644393	
LOCUS	AZ644393 496 bp DNA linear GSS 14-DEC-2000
DEFINITION	IM0508M1F Mouse 10kb plasmid UUC1M library Mus musculus genomic clone UUCGIM0508M18 F, DNA sequence.
ACCESSION	AZ644393
VERSION	AZ644393.1 GI:11772878
KEYWORDS	GSS.
SOURCE	house mouse.
ORGANISM	Mus musculus

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
Eumariota; Chorizota; Chordata; Crinata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciuromathia; Muridae; Murinae; Mus.	1 (bases 1 to 496)			
Dunn D., Aoyagi A., Barber M., Baocorn T., Duval B., Hanll C., Islam H., Longacre S., Mahmoud M., Meenen F., Pedersen T., Reilly M., Rose M., Rose R., Stokes R., Tinney A., von Niederhausern A. and Wright D., Weiss R.		Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts		Unpublished (2000)
				Contact: Robert B. Weiss

University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
84112, USA
Tel.: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0508 row: M column: 18
Seq primer: CGTTGTAACAACGACGGCAGT
Class: plasmid ends
High quality sequence stop: 496.
Location/Qualifiers
1. 496

```

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="U06C1M0508M18"
/clone_id="Mouse 10kb plasmid U06C1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
/notes="Vector: pMD42ny; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were

```

ligated to the blunt ends in high molar excess. The
 adaptor DNA was purified and size-selected for a 9.5 to
 10.5 kb range using preparative agarose gel
 electrophoresis. Vector DNA was prepared from a derivative
 of pMD42 (g1147311419b1aF129072.1), a copy-number
 inducible derivative of plasmid R1. The vector was ligated
 with adaptors complementary to the insert adaptors and
 purified. The sheared, adaptor mouse DNA was annealed to
 adaptor vector DNA, and transformed into
 chemically-competent *E. coli* XL10-Gold (Stratagene) cells
 and selected for ampicillin resistance."

Query Match	Similarity	Score	DB	Length
Best Local Match	86.5%	Pred. No. 1.6e-32;		
Matches 180;	Conservative	0;	Mismatches 28;	Indels 0;
				Gaps 0;
QY	293	ccatcgttagtcatcttctagacgatlbaatctgagacacgtcttcagaaatttctiaacagac	352	
Db	2	ccatcagtagatgcatcttctttagatgatgaactgaactgaactgacgtcttcagaaatttctiaacagat	61	
QY	353	attgtacacaaaaccttggagggtttgtcctttaaanaattagatgcatctatacaaatccg	412	
Db	62	attgtgacaaaaaaaccttggagggtttgtcctttaaanaactagatgcatctatacaaatccg	121	
QY	413	cttataaaaaatatatcatcaccatcatccaagtgctgttttcagataatggagaag	472	
Db	122	cttgttttataaaatacatattcatctccacacgtcccgaaatgctgatttttcagataatggagaag	181	
QY	473	atgcacattgcagaatctgttatacaaa	500	
Db	182	atgaccttgcagaatgctgttatacaaa	209	

RESULT	6
BH110525/c	
LOCUS	BH110525
DEFINITION	717 bp DNA linear GSS 19-JUN-2001 Rpci-24-340G10.TJ Rpci-24 Mus musculus genomic clone Rpci-24-340G10 , DNA sequence.
ACCESSION	BH110525
VERSION	BH110525
KEYWORDS	BH110525.1 GI:14944731
SOURCE	GSS.
ORGANISM	house mouse. Mus musculus
REFERENCE	Euaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murine; Mus. 1 (bases 1 to 717)
AUTHORS	Zhao,S., Nieman,W., Malek,J., Shatsman,S., Akincet,B., Levins,M., Tasagay,G., Geer,K., Krol,M., Shwartsbyan,A., Gebregeorgis,E., Russell,D., de Jong,P. and Fraser,C.M.
TITLE	Mouse BAC End Sequences from Library Rpci-24
JOURNAL	Unpublished (1999)
COMMENT	Other_GSSs: Rpci-24-340G10.TV

FEATURES
source
 Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel.: 301 838 0200
 Fax: 301 838 0208
 Email: szhao@tigr.org
 Clones are derived from the mouse BAC library RPc1-24. For BAC
 library availability, please contact Pieter de Jong
 (pejong@nail.chi.org). Clones may be purchased from BACPAC
 Resources (<http://www.chori.org/bacpac/orderingframe.htm>). BAC end
 page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
 plate: 340 row: G column: 10
 Seq primer: SP6
 Class: BAC ends.
 Location/Qualifiers
 1..717
 /organism="Mus musculus"

BASE COUNT	206 a	125 c	133 g	253 t
ORIGIN				
Query Match	31.8%, Score 159.2; DB 12; Length 717;			
Best Local Similarity	84.5%, Pred. No. 2e-31;			
Matches 191:	Conservative	0;	Mismatches 33;	Indels 2; Gaps 1;
OY	277	catgactcaagatccatcgtagtcattttagaagatgaatcgaagacagcttc	336	
Db	717	CATCAGACTCAGAGATCCCATCAGTACGTATTTTAATGATGAAGACTGACCTTCC	658	
OY	337	agaatttttagagacattgttacaacacttggaggggtttgtccttaaaaattagaagc	396	
Db	657	AGAAATCTTAGAGATTTGATGACAAAACCTGGAGGGATGTGCTGAAAAGACTGTGAATAC	598	
OY	397	actctacaacatccgcttatataaaat--atctatctaccattaccagatgctgt	454	
Db	597	CTCTATTGACGATCCACCTGTTTAAAAATACACATCATTTCCCCACCTCCGAGTCTATT	538	
OY	455	ttgcagataatggagaagatgccattgcagaaattgtgtaacaa	500	
Db	537	TTGCAGATATGAGAGATACCTCTCAGAGAGTTGTGTAATCAAA	492	
RESULT 7				
CNS04HOG/c	883 bp DNA linear GSS 21-MAY-2000			
LOCUS	CNS04HOG			
DEFINITION	Tetraodon nigroviridis genome survey sequence T7 end of clone 110L02 of library G from Tetraodon nigroviridis, genomic survey sequence.			
ACCESSION	AL291193			
VERSION	AL291193.1 GI:8029773			
KEYWORDS	GSS; genome survey sequence.			
SOURCE	Tetraodon nigroviridis.			
ORGANISM	Tetraodon nigroviridis.			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.			
AUTHORS	1 (bases 1 to 883) Roest-Crollius,H., Jallou,O., Dasilva,C., Fizeses,C., Fisher,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.			
TITLE	Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis			
JOURNAL	Unpublished			
REFERENCE	2 (bases 1 to 883) Roest-Crollius,H., Jallou,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizeses,C., Wincker,P., Brothier,P., Quetier,F., Saurin,W. and Weissenbach,J.			
AUTHORS	Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence			
TITLE	Unpublished			
JOURNAL	3 (bases 1 to 883)			
REFERENCE	Genoscope.			
AUTHORS	Direct Submision			
TITLE	Submitted (12-APR-2000) to the EMBL/Genbank/DBJ databases			
JOURNAL	This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at			
COMMENT	http://www.genoscope.cns.fr/Tetraodon.			

FEATURES		Location/Qualifiers
Source	1..883	
	/organism="Tetradon nigriviridis"	
	/db_xref="taxon:99883"	
	/clone="110L02"	
	/clone_lib="C"	
	/note="Genoscope sequence ID : CDBG110DP01LPI-end : T7"	
BASE COUNT	233 a 227 c 197 g 223 t	3 others
ORIGIN		
Query Match	28.4%	Score 142; DB 12; Length 883;
Best Local Similarity	59.8%	Pred. NO. 7e-27;
Matches 238; Conservative	0; Mismatches 160; Indels 0; Gaps 0;	
Oy	8 acagaagaacacatctacacagctcagctcagctcctaaatccagaacagatcttcacgtcttacc 67	
Db	398 ACAGGGAACACCTTGTACTCAGCTGCAATGCTGAGCCCGGAGCGGTCCGCTCGACTCATC 339	
Oy	68 aagaagaatgaatgaatcacatctctgcccgcgcgcagagaatgatgtgtcattggtatccattt 127	
Db	338 AAGGACGCTCTGCTCTCCTCAGCGGCGCCACAGAGGAGCTTCGCTGTGATGAGGAGCGCTGGA 279	
Oy	128 gatgaacaacagagaatccaccatctgtttcattcagcttaagaatggtgttgtaaaattcttat 187	
Db	278 AACAGAGAGCTGAGACATCCACCCCTTTCGCGCTCAGACAGATGCTTGGAAACCATCTCTAC 219	
Oy	188 atacattttcagaagatgtgaatttatattgatagatgcacattaccccaagaactata 247	
Db	218 ATACATTTTGGCTGAGAGGACTTGACACTTTTCGAGACATGCTTTGATCCCACTCGTGCCA 159	
Oy	248 ctagaagaaggttcagacatgctgtggaactcatctagaactcagaatccatcgtctagtcatt 307	
Db	158 CTTGAGGAAACATGACACGAGATTGATCTTGCGCGCTCAGAACATCTTCAACCCATCATTA 99	
Oy	308 ttagaacatgaatctcgaagcacaagcttccagaatttttaagcagacatgttcaaaaactt 367	
Db	98 TTTGGCGATCGCGAAGAAAGCAGACCCCTTCGGAACCTTCGAGACATCATGTGAAAAAGCTC 39	
Oy	368 ggaagggttctccctaaaaatgaatgacatctata 405	
Db	38 GGAGGAGCGGTGATGACAAAGCTGATTCGTTTGCA 1	
RESULT 8		
LOCUS	BB561400	202 bp mRNA linear EST 01-AUG-2000
DEFINITION	BB561400 RIKEN full-length enriched, 10 days neonate olfactory	
ACCESSION	BB561400	brain Mus musculus cDNA E530213Jf07 3', mRNA sequence.
VERSION	BB561400.1	GI:9647766
KEYWORDS	EST.	
SOURCE	house mouse.	
ORGANISM	Mus musculus	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
	1 (bases 1 to 202)	
	Komno H., Aitawa K., Akahira S., Akiyama J., Arahawa T., Carninci	
	P., Endo T., Fukuda S., Fukunishi Y., Hara A., Hayatsu N.,	
	Hirozane T., Hori F., Ishii Y., Ishikawa J., Ishikawa T., Itoh M.,	
	Iizawa M., Kadota K., Kagawa I., Kai C., Kawai J., Kikuchi N.,	
	Kiyosawa H., Kojima Y., Kondo S., Koya S., Kurihara C., Kusakabe M.,	
	Matsuyama T., Miki R., Mizuno Y., Nakamura M., Oda H., Okazaki Y.,	
	Ono T., Owa C., Saito H., Sakai C., Sato K., Shibata K., Shibata	
	Y., Shigemoto Y., Shinagawa A., Shiraki T., Sogabe Y., Sugahara Y.,	
	Suzuki H., Suzuki H., Tagawa A., Takahashi F., Tomimasa N., Toya	
	T., Tsunoda Y., Watabiki A., Watanabe S., Yamamura T., Yamanaoka I.,	
	Yano R., Yasunishi A., Yokota T., Yoshida K., Yoshiki A., Yoshino	
	M., Muramatsu M. and Hayashizaki Y.	
	RIKEN Mouse ESTs (Komno H., et al.)	
	Unpublished (2000)	
TITLE		
JOURNAL		
COMMENT		
	Contact: Yoshihide Hayashizaki	
	Laboratory for Genome Exploration Research Group, RIKEN Genomic	

VERSION	BF32169.1	GI:11296947
KEYWORDS	EST.	
SOURCE	human.	
ORGANISM	Homo sapiens	
	Eumetazoa; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
REFERENCE	1 (bases 1 to 266)	
AUTHORS	Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsumura, A., Bata, G.S., Simpson, D.H., Brunsfeld, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.	
TITLE	Shotgun sequencing of the human transcriptome with ORF expressed sequence tags	
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)	
MEDLINE	20202663	
COMMENT	Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-27047001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=MR06t2-MR0-AN0083-160900-003-d10&t3=2000-09-16&t4=1) Seq primer: puc 18 forward High quality sequence stop: 11. Location/Qualifiers 1..266 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_lib="AN0083" /dev_stage="Adult" /note="Organ: amnion_normal; Vector: puc18; Site:1: Sma1; Site_2: Sma1; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - ludwig institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions." BASE COUNT 71 a 67 c 52 g 76 t ORIGIN	
Query Match	11.4%;	Score 57; DB 10; Length 266;
Best Local Similarity	100.0%;	Pred. No. 0.00012;
Matches 57; Conservative 0;	Mismatches 0;	Indels 0; Gaps 0;
Y	10	aggaagaccatgctactcagctgcagctcttaattcagaagattgcagctctat 66 Db 210 AGGAGACCACTACTCAGCTGACGCTTCAATCCAGACGATTTCACGCTTAT 266
RESULT 11		
LOCUS	BH571801	752 bp DNA linear GSS 14-DEC-2001
DEFINITION	BOHLS54TR BOHL Brassica oleracea genomic clone BOHLS54, DNA sequence.	
ACCESSION	BH571801	
VERSION	BH571801.1	GI:17823640
KEYWORDS	GSS.	
SOURCE	Brassica oleracea.	
ORGANISM	Brassica oleracea	
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eucosids II; Brassicales; Brassicaceae; Brassica.	
REFERENCE	1 (bases 1 to 752)	
AUTHORS	Town, C.D., Van Aken, S., Uterback, T. and Fraser, C.M.	
TITLE	Whole genome shotgun sequencing of Brassica oleracea	

JOURNAL Unpublished (2001)
COMMENT Other_GSSs: BOHLS54TF
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid produced by Tom Osborn.
Seq primer: TR
Class: sheared ends.

FEATURES
Source Location/Qualifiers
1..752
/organism="Brassica oleracea"
/strain="TO100DH3"
/db_xref="taxon:3712"
/clone="BOHLS54"
/clone_lib="BOHL"
/vector: phosI; site_1: BstXI; 2-3 kb sheared genomic DNA inserted into phosI using BstXI linkers"

BASE COUNT 267 a 114 c 109 g 262 t

ORIGIN

```

Query Match      8.8%; Score 44; DB 12; Length 752;
Best Local Similarity 45.1%; Pred. No. 0.41;
Matches 164; Conservative 0; Mismatches 200; Indels 0; Gaps 0;

OY   124 attgatgaaacagagaatcaccccatcgttcctcaatgcgttaagatgcttggaataact 183
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db    466 ATGACATGGCTACTGAAATTGTGANTGACAGTCCTTAGCTTAATATGACATGTACAAATGC 407

OY   184 ttatatcatctttccagaggatttbaactttatgatgagatggccaacttaccccagac 243
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db    406 ATTATATGTTAATTTATATATTTTTTGCTAACACTTTTAAATAATGGCAATAATTCATATATT 347

OY   244 tatactagaggagagglcagacatgltggaactcatlagacctcaggaattccatcgtagt 303
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db    346 ACATTATATGTCGAATTTATATTTTTGGAAATTTTTTGAATGCGAATCGCAATCATTAATCAT 287

OY   304 cattctagacgatgaactctgaagcaagctccagaatttttagcagacatctgtacaana 363
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db    286 CATTTAAATTAATATATATTTCAATATGACATATATAAATTCGAAATATATTAATTATATAT 227

OY   364 acttgagagggttgcctctaanaaatgatgacatctacaacatccgcctattaaaaa 423
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db    226 ATTTTAAATTAATTAACAATTTTATTAACGTAAAAATTTCCAAAAATGCCATACAAATTTTATGA 167

OY   424 atatactaccatcacatcaccaagtgcgtttcttcagataaaygagaaagatgccatlgca 483
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db    166 AAATTAATAAATTTTAATGCTAAATCATTATTTTCTCTGTTGACGGCGGACGATTAAA 107

OY   484 gaaga 487
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db    106 GAAA 103
```

RESULT 12
A2066617/c LOCUS A2066617 440 bp DNA linear GSS 30-MAR-2000
DEFINITION RPCR-23-433L3.rv RPCR-23 Mus musculus genomic clone RPCR-23-433L3,
DNA sequence.
ACCESSION A2066617
VERSION A2066617.1 GI:7357869
KEYWORDS GSS.
ORGANISM house mouse.
SOURCE Mus musculus
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathii; Muridae; Murinae; Mus.
Zhao,S., Nierman,W., Feildblum,T., Malek,J., Shatsman,S., Akirret
,B., Levins,M., McGann,S., Geer,K., Kroi,M., de Jong,P.,
and Fraser,C.M.

TITLE Mouse BAC End Sequences from library RPCI-23
JOURNAL Unpublished (1999)
COMMENT Other GSSs: RPCI-23-433L3.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@edlong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>) or from Resea ch Genetics (info@resgen.com). BAC end page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 433 row: L column: 3
Seq primer: 77
Class: BAC ends.
Location/Qualifiers

FEATURES
source 1..440
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-433L3"
/clone_lib="RPCI-23"
/sex="Female"
/lab_host="DH10B"
/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT 120 a 74 c 73 g 173 t
ORIGIN

Query Match 8.8%; Score 43.8; DB 12; Length 440;
Best Local Similarity 87.3%; Pred. No. 0.4;
Matches 48; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Oy 446 agtgcgtcttgcagataatggaagatgcacatgcagaatctgtaatacaaa 500
|||||
Db 418 AGTGATATTTCAGATATGAGAGATACCTCTACAGAGTGTGTATCAAA 364

RESULT 13
BH131081 895 bp DNA linear GSS 07-AUG-2001
LOCUS BH131081/C
DEFINITION ENTMOPI6TF Entamoeba histolytica sheared DNA Entamoeba histolytica genomic DNA sequence.
ACCESSION BH131081
VERSION BH131081.1 GI:15089550
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica.
Eukaryota; Entamoebidae; Entamoeba.
1 (bases 1 to 895)
Loftus, B., Wang, Z., Van Aken, S. and Fraser, C.
Determination of clone end sequences from Entamoeba histolytica HMI:IMSS sheared DNA library (2001)
Unpublished (2001)
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: b.loftus@tigr.org
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared DNA library

Seq primer: M13-Forward
Class: shotgun
High quality sequence start: 43
High quality sequence stop: 830.
Location/Qualifiers

FEATURES
source 1..895
/organism="Entamoeba histolytica"
/strain="HMI:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/note="Vector: pBACe3.6; Site_1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a light size distribution (~2 kb). The v + 1 method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaubin and B. Barrell, Oxford University Press, 1999)."

BASE COUNT 302 a 117 c 103 g 373 t
ORIGIN

Query Match 8.3%; Score 41.4; DB 12; Length 895;
Best Local Similarity 50.2%; Pred. No. 2.1;
Matches 102; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

Oy 32 cagctcttaatacagaacagattgcacgtctatacagaagaatgaatacatctcg 91
|||
Db 324 CAATTTGTTAACAAAGAACAAATGATGGAAGAAACAGAAAGAGAAATATTTAT 265
Oy 92 cctgcagagaaatgttcaatgtatccatttgatgaagaacagaatccaccat 151
|||
Db 264 ACTTCATTAGTACTGTTGTTATTCCTTACCTTAATCAAGAAATGCTTGAATTA 205
Oy 152 gttctgccttaagatggttggaagaatcttatatacatcttcagagagattcg 211
|||
Db 204 TTTCAAAATTAATTAATTCCTTTTAAACAGATTATTTTAAATTTCTAGACAGACATACT 145
Oy 212 ttaattgatgagatgcacttat 234
|||
Db 144 ACAATTAAATTAATAAGTTACTTAT 122

RESULT 14
BH136596 896 bp DNA linear GSS 07-AUG-2001
LOCUS BH136596
DEFINITION ENTMO65TF Entamoeba histolytica sheared DNA Entamoeba histolytica genomic DNA sequence.
ACCESSION BH136596
VERSION BH136596.1 GI:15095657
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica.
Eukaryota; Entamoebidae; Entamoeba.
1 (bases 1 to 896)
Loftus, B., Wang, Z., Van Aken, S. and Fraser, C.
Determination of clone end sequences from Entamoeba histolytica HMI:IMSS sheared DNA library (2001)
Unpublished (2001)
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: b.loftus@tigr.org
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared DNA library

Seq primer: M13-Forward
Class: shotgun
High quality sequence start: 47
High quality sequence stop: 723.
Location/Qualifiers
1. 896

Seq primer: M13-Forward
Class: shotgun
High quality sequence start: 299
High quality sequence stop: 873.
Location/Qualifiers
1. 940

FEATURES
source
/organism="Entamoeba histolytica"
/strain="HMI:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/note="Vector: PHOS1; Site 1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaubin and B. Barrell, Oxford University Press, 1999)."

FEATURES
source
/organism="Entamoeba histolytica"
/strain="HMI:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/note="Vector: PHOS1; Site 1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaubin and B. Barrell, Oxford University Press, 1999)."

Query Match 8.3%; Score 41.4; DB 12; Length 896;
Best Local Similarity 50.2%; Pred. No. 2.1;
Matches 102; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

Query Match 8.3%; Score 41.4; DB 12; Length 940;
Best Local Similarity 50.2%; Pred. No. 2.1;
Matches 102; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

BASE COUNT 402 a 70 c 130 g 294 t
ORIGIN
QY 32 cagcttctaataccagaagatttgcagcgtctatcagaagaatgaatcattctcg 91
DB 154 CAATTTGTTAAACAAGACAATGATGAGAAACAAGTAAAGAAAGAATTTTAT 213
QY 92 cctgcagagaattgattcaatgltatccattgataaacaagaatccaccatc 151
DB 214 ACTTCATTAGTACTTATTTGTTATTTATCTTCACCTAATTCAAGAAATGCTGAATTA 273
QY 152 gtttcattgcttaagaatggttggaaaatccttatatacatcttccaaggattgact 211
DB 274 TTACAAATTAATAATCTCTTTTAAACAGTATTTTAAATTCYAGAAACATACACT 333
QY 212 ttattgatgagatgcacattat 234
DB 334 ACAATTATATAAAGTTACTTAT 356

BASE COUNT 393 a 92 c 149 g 306 t
ORIGIN
QY 32 cagcttctaataccagaagatttgcagcgtctatcagaagaatgaatcattctcg 91
DB 483 CAATTTGTTAAACAAGACAATGATGAGAAACAAGTAAAGAAAGAATTTTAT 542
QY 92 cctgcagagaattgattcaatgltatccattgataaacaagaatccaccatc 151
DB 543 ACTTCATTAGTACTTATTTGTTATTTATCTTCACCTAATTCAAGAAATGCTGAATTA 602
QY 152 gtttcattgcttaagaatggttggaaaatccttatatacatcttccaaggattgact 211
DB 603 TTACAAATTAATAATCTCTTTTAAACAGTATTTTAAATTCYAGAAACATACACT 662
QY 212 ttattgatgagatgcacattat 234
DB 663 ACAATTATATAAAGTTACTTAT 685

RESULT 15
A2547902 940 bp DNA linear GSS 14-NOV-2000
DEFINITION ENTDP1ITF Entamoeba histolytica Sheared DNA Entamoeba histolytica
genomic, DNA sequence.
ACCESSION A2547902
VERSION A2547902.1 GI:11170987
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica.
REFERENCE 1 (bases 1 to 940)
AUTHORS Loftus,B., Van Aken,S. and Fraser,C.
TITLE Determination of clone end sequences from Entamoeba histolytica
JOURNAL HMI:IMSS sheared DNA library
COMMENT Unpublished (2000)
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjlloftus@tigr.org
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
DNA library

Search completed: May 22, 2002, 05:31:11
Job time: 4071 sec

.

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 22, 2002, 06:59:14 ; Search time 3530.57 Seconds
(without alignments)
118.545 Million cell updates/sec

Title: US-09-693-205-47

Perfect score: 20
Sequence: 1 gtgatgcacacttgcact 20

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*
1: gb.ba:*
2: gb.htg:*
3: gb.in:*
4: gb.om:*
5: gb.ov:*
6: gb.pat:*
7: gb.ph:*
8: gb.pl:*
9: gb.pr:*
10: gb.ro:*
11: gb.sts:*
12: gb.sy:*
13: gb.un:*
14: gb.vl:*
15: em.ba:*
16: em.fun:*
17: em.hum:*
18: em.in:*
19: em.mu:*
20: em.om:*
21: em.or:*
22: em.ov:*
23: em.pat:*
24: em.ph:*
25: em.pl:*
26: em.ro:*
27: em.sts:*
28: em.un:*
29: em.vl:*
30: em.htg.hum:*
31: em.htg.inv:*
32: em.htg.other:*
33: em.htgo.inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result Query
No. Score Match Length DB ID Description

1	20	100.0	12793	6	AX119931	Sequence
2	20	100.0	12793	9	AF193556	AF193556 Homo sapi
3	20	100.0	92693	9	AL157766	AL157766 Human DNA
4	20	100.0	99819	2	AC079761	AC079761 Homo sapi
5	18.4	92.0	11492	6	AX119933	AX119933 Sequence
6	18.4	92.0	11493	10	AF193557	AF193557 Mus muscu
7	17.4	87.0	434	11	G37224	G37224 SHGC-57260
8	17.4	87.0	766	3	AF113358	AF113358 Ips latid
9	17.4	87.0	766	3	AF113359	AF113359 Ips latid
10	17.4	87.0	175309	2	AC069021	AC069021 Homo sapi
11	17.4	87.0	176713	9	AP000751	AP000751 Homo sapi
12	17.4	87.0	192913	2	AP001265	AP001265 Homo sapi
13	17.4	87.0	216521	9	AC073587	AC073587 Homo sapi
14	17.4	87.0	241382	2	AC019279	AC019279 Homo sapi
15	17	85.0	993	3	P095073	P095073
16	17	85.0	8028	3	MIPOND2	MIPOND2
17	17	85.0	69648	2	AC099825	AC099825 Papio ham
18	17	85.0	129504	9	AL445469	AL445469 Human DNA
19	17	85.0	160866	2	AC092532	AC092532 Papio cyn
20	17	85.0	174897	2	AC090965	AC090965 Papio cyn
21	17	85.0	178317	9	AC012081	AC012081 Homo sapi
22	17	85.0	201957	2	AC026359	AC026359 Homo sapi
23	16.8	84.0	2544	2	AC020530	AC020530 Drosophila
24	16.8	84.0	31887	2	AL450999	AL450999 Human DNA
25	16.8	84.0	61394	2	AC102771	AC102771 Mus muscu
26	16.8	84.0	68997	2	DMB6F4	DMB6F4
27	16.8	84.0	71999	2	AC101210	AC101210 Mus muscu
28	16.8	84.0	72648	2	AC101171	AC101171 Mus muscu
29	16.8	84.0	72648	2	AC101171	AC101171 Mus muscu
30	16.8	84.0	84894	8	AP003251	AP003251 Oryza sat
31	16.8	84.0	118380	2	AC094213	AC094213 Rattus no
32	16.8	84.0	123943	10	AL592547	AL592547 Mouse DNA
33	16.8	84.0	130084	2	AC092265	AC092265 Homo sapi
34	16.8	84.0	152042	9	AL137857	AL137857 Human DNA
35	16.8	84.0	166704	2	AC078821	AC078821 Homo sapi
36	16.8	84.0	169576	2	AL663043	AL663043 Mus muscu
37	16.8	84.0	170669	9	AC007907	AC007907 Homo sapi
38	16.8	84.0	171386	2	AC079465	AC079465 Homo sapi
39	16.8	84.0	173525	2	AC067893	AC067893 Homo sapi
40	16.8	84.0	182004	2	AC097609	AC097609 Rattus no
41	16.8	84.0	184649	2	AC025185	AC025185 Homo sapi
42	16.8	84.0	188553	2	AC093394	AC093394 Bos tauru
43	16.8	84.0	210730	2	AC015975	AC015975 Homo sapi
44	16.8	84.0	216016	2	AL626775	AL626775 Mus muscu
45	16.8	84.0	228350	2	AC093396	AC093396 Bos tauru

ALIGNMENTS

RESULT 1
AX119931
LOCUS AX119931 12793 bp DNA linear PAT 11-MAY-2001
DEFINITION Sequence 1 from Patent WO0129266.
ACCESSION AX119931
VERSION AX119931.1 GI:14036678
KEYWORDS
SOURCE
ORGANISM Homo sapiens
human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 12793)
Hudson,T.J., Engert,J. and Richter,A.
Identification of arsacs mutations and methods of use therefor
Patent: WO 0129266-A 1 26-APR-2001;
JOURNAL MCGILL UNIVERSITY (CA) : Hospital Sainte-Justine (CA)
FEATURES
source
1..12793
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT 4163 a 2256 c 2487 g 3887 t
ORIGIN

AUTHORS
JOURNAL

COMMENT

Tromans, A.
Submitted (11-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humpuety@sanger.ac.uk
Requests: clonequest@sanger.ac.uk
On April 12, 2001 this sequence version replaced g1:12709668.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em, EMBL; Sw, S, SWISSPROT; Tr, TrEMBL; Wp, WormPEP; Information on the WormPEP
database can be found at
http://www.sanger.ac.uk/projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 13, constructed by the Sanger Centre Chromosome 13
Mapping Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr13>
RP11-40020 is from the library RPCI-11.1 constructed by the group
of Pieter de Jong. For further details see
<http://www.chori.org/dacpac/home.htm>
VECTOR: pBAC3.6

FEATURES

source

1. 92693
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="13"
/map="g12.11-12.2"
/clone="RP11-40020"
/clone_1lb="RPCI-11.1"
2390..2485
/note="MIR repeat: matches 81..192 of consensus"
2562..2673
/note="MIR repeat: matches 68..183 of consensus"
3896..4201
/note="AluY repeat: matches 3..308 of consensus"
5122..5397
/note="MER46C repeat: matches 1..286 of consensus"
18986..19294
/note="AluY repeat: matches 1..310 of consensus"
19644..19873
/note="MER46A repeat: matches 1..235 of consensus"
20613..20912
/note="AluSg1 repeat: matches 1..300 of consensus"
23342..23651
/note="AluSg1 repeat: matches 1..309 of consensus"
24769..24891
/note="L2 repeat: matches 2554..2682 of consensus"
23871..26011
/note="L2 repeat: matches 2356..2495 of consensus"
26033..26109
/note="L2 repeat: matches 2601..2688 of consensus"
26245..26344
/note="L2 repeat: matches 2154..2255 of consensus"
26338..27096
/note="MIR repeat: matches 3..175 of consensus"
27150..27653
/note="L2 repeat: matches 1063..1644 of consensus"

repeat_region 28522..28891
/note="THE1B repeat: matches 1..364 of consensus"
repeat_region 29447..29834
/note="LIME3A repeat: matches 5787..6164 of consensus"
repeat_region 36098..36415
/note="AluX repeat: matches 1..308 of consensus"
repeat_region 37202..37414
/note="MIR repeat: matches 22..262 of consensus"
repeat_region 37963..38254
/note="AluSg repeat: matches 9..301 of consensus"
repeat_region 38703..39008
/note="AluSg repeat: matches 1..306 of consensus"
repeat_region 39790..40093
/note="AluSg repeat: matches 1..304 of consensus"
repeat_region 40126..40416
/note="AluSg repeat: matches 1..292 of consensus"
repeat_region 40444..40733
/note="AluSg repeat: matches 1..292 of consensus"
41322..41405
/note="Single clone region. Assembly confirmed by
restriction digest data."
repeat_region 41341..41788
/note="AluSg repeat: matches 1..248 of consensus"
repeat_region 44790..45101
/note="AluSg repeat: matches 1..313 of consensus"
repeat_region 45261..45312
/note="13 copies 4 mer tgt 888 conserved"
repeat_region 45899..46206
/note="AluY repeat: matches 1..307 of consensus"
repeat_region 46754..47052
/note="AluY repeat: matches 1..298 of consensus"
repeat_region 47067..47365
/note="AluY repeat: matches 1..299 of consensus"
repeat_region 47477..47873
/note="L1MA10 repeat: matches 5950..6322 of consensus"
repeat_region 47889..48229
/note="AluX repeat: matches 1..312 of consensus"
repeat_region 49168..49212
/note="Alu repeat: matches 85..126 of consensus"
repeat_region 49620..49693
/note="L2 repeat: matches 1685..1757 of consensus"
repeat_region 50704..51032
/note="AluX repeat: matches 1..308 of consensus"
52204..53009
/note="CpG Island"
/note="evidence-not-experimental"
repeat_region 53978..54137
/note="AluX repeat: matches 1..160 of consensus"
54179..54511
/note="L1MB6 repeat: matches 5822..6172 of consensus"
55685..55949
/note="AluX repeat: matches 1..305 of consensus"
repeat_region 57331..57390
/note="30 copies 2 mer ga 75% conserved"
repeat_region 57357..57392
/note="9 copies 4 mer gaga 91% conserved"
repeat_region 57753..57930
/note="MIR repeat: matches 82..262 of consensus"
repeat_region 58260..58389
/note="MIR repeat: matches 2..153 of consensus"
58564..58611
/note="24 copies 2 mer ca 93% conserved"
repeat_region 59350..59533
/note="AluSg repeat: matches 129..313 of consensus"
repeat_region 59992..60223
/note="AluSg repeat: matches 85..299 of consensus"
repeat_region 61036..61144
/note="L2 repeat: matches 2581..2696 of consensus"
repeat_region 62008..62187
/note="TIGER1 repeat: matches 2238..2418 of consensus"
repeat_region 62188..62316
/note="AluY repeat: matches 1..129 of consensus"
repeat_region 62330..62363

```
repeat_region /note="Alu repeat: matches 261. .294 of consensus"
62362. .62565
/note="TIGER1 repeat: matches 1586. .1787 of consensus"
repeat_region 62566. .62865
/note="AluSg repeat: matches 1. .302 of consensus"
repeat_region 62866. .64385
/note="TIGER1 repeat: matches 46. .1586 of consensus"
repeat_region 64386. .64694
/note="AluY repeat: matches 1. .306 of consensus"
repeat_region 64695. .64713
/note="TIGER1 repeat: matches 29. .46 of consensus"
repeat_region 65068. .65395
/note="L1PB2 repeat: matches 5405. .5733 of consensus"
repeat_region 65396. .65569
/note="AluY repeat: matches 136. .309 of consensus"
repeat_region 65571. .65640
/note="L1PB2 repeat: matches 5728. .5791 of consensus"
repeat_region 65696. .65717
/note="11 copies 2 mer ta 100% conserved"
repeat_region 65725. .66096
/note="L1PB2 repeat: matches 5789. .6155 of consensus"
repeat_region 66371. .66410
/note="10 copies 4 mer tctg 82% conserved"
repeat_region 67586. .67886
/note="AluY repeat: matches 1. .299 of consensus"
repeat_region 69748. .69930
/note="MIR repeat: matches 6. .248 of consensus"
repeat_region 70957. .71267
/note="AluY repeat: matches 1. .311 of consensus"
repeat_region 71279. .71413
/note="MER21B repeat: matches 548. .680 of consensus"
repeat_region 71411. .71737
/note="MER31A repeat: matches 47. .485 of consensus"
repeat_region 71780. .72075
/note="AluSx repeat: matches 1. .295 of consensus"
repeat_region 72143. .72256
/note="MER31-internal repeat: matches 42. .175 of
consensus"
repeat_region 72454. .72865
/note="MER31-internal repeat: matches 332. .739 of
consensus"
repeat_region 72873. .73249
/note="MER31-internal repeat: matches 883. .1261 of
consensus"

Query Match 100.0% Score 20; DB 9; Length 92693;
Best Local Similarity 100.0% Pred. No. 3;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gtaatggcacttgact 20
Db 12437 GTCAATGCCACCTTGCAC 12418
```

```
RESULT 4
AC079761 99819 bp DNA linear HTG 10-SEP-2000
LOCUS AC079761.c
DEFINITION Homo sapiens chromosome UNK clone RP11-143G17, *** SEQUENCING IN
ACCESSION AC079761
VERSION AC079761.1 GI:10047966
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 99819)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 99819)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (10-SEP-2000) Genome Sequencing Center, Washington
```

COMMENT

```
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Web site: http://genome.wustl.edu/gsc/index.shtml
Project Information -----

* NOTE: This is a 'working draft' sequence. It currently
* consists of 44 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1137: contig of 1137 bp in length
1138 1237: gap of unknown length
1238 2538: contig of 1301 bp in length
2539 2638: gap of unknown length
2639 3976: contig of 1338 bp in length
3977 4076: gap of unknown length
4077 5355: contig of 1279 bp in length
5356 5455: gap of unknown length
5456 6757: contig of 1302 bp in length
6758 6857: gap of unknown length
6858 8570: contig of 1713 bp in length
8571 9925: gap of unknown length
9926 10025: contig of 1255 bp in length
10026 11426: contig of 1401 bp in length
11427 11526: gap of unknown length
11527 13266: contig of 1740 bp in length
13267 13366: gap of unknown length
13367 14794: contig of 1428 bp in length
14795 14894: gap of unknown length
14895 16054: contig of 1160 bp in length
16055 16154: gap of unknown length
16155 17395: contig of 1241 bp in length
17396 17495: gap of unknown length
17496 19287: gap of 1792 bp in length
19288 19387: gap of unknown length
19388 21294: contig of 1907 bp in length
21295 21394: gap of unknown length
21395 22944: contig of 1550 bp in length
22945 23044: gap of unknown length
23045 24421: contig of 1377 bp in length
24422 24521: gap of unknown length
24522 25870: contig of 1349 bp in length
25871 25970: gap of unknown length
25971 27230: contig of 1260 bp in length
27231 27330: gap of unknown length
27331 28778: contig of 1448 bp in length
28779 28878: gap of unknown length
28879 30893: contig of 2015 bp in length
30894 32460: gap of unknown length
32461 32560: gap of unknown length
32561 33984: contig of 1424 bp in length
33985 34084: gap of unknown length
34085 35285: contig of 1201 bp in length
35286 35385: gap of unknown length
35386 37184: contig of 1799 bp in length
37185 37284: gap of unknown length
37285 39172: gap of 1888 bp in length
39173 39272: gap of unknown length
39273 40874: contig of 1602 bp in length
40875 42893: contig of 1919 bp in length
42894 42993: gap of unknown length
42994 44384: contig of 1391 bp in length
44385 44484: gap of unknown length
44485 45999: contig of 1515 bp in length
```

```
* 46000 46099: gap of unknown length
* 46100 48669: contig of 2570 bp in length
* 48670 48670 48669: gap of unknown length
* 48770 50798: contig of 2029 bp in length
* 50799 50898: gap of unknown length
* 50899 52809: contig of 1911 bp in length
* 52810 52909: gap of unknown length
* 52910 55127: contig of 2218 bp in length
* 55128 55227: gap of unknown length
* 55228 58087: contig of 2860 bp in length
* 58088 58187: gap of unknown length
* 58188 61004: contig of 2817 bp in length
* 61005 61104: gap of unknown length
* 61105 64185: contig of 3081 bp in length
* 64186 64285: gap of unknown length
* 64286 67105: contig of 2820 bp in length
* 67106 67205: gap of unknown length
* 67206 70837: contig of 3632 bp in length
* 70838 70937: gap of unknown length
* 70938 75937: contig of 4900 bp in length
* 75938 80452: gap of unknown length
* 80453 80552: gap of unknown length
* 80553 84661: contig of 4109 bp in length
* 84662 84761: gap of unknown length
* 84762 90542: contig of 5781 bp in length
* 90543 90642: gap of unknown length
* 90643 94348: contig of 3706 bp in length
* 94349 94449 99819: contig of 5371 bp in length.
```

FEATURES
source

```
misc_feature /note="assembly_name:Contig20"  
misc_feature 1238. 2538 /note="assembly_name:Contig23"  
misc_feature 2639. 3976 /note="assembly_name:Contig30"  
misc_feature 4077. 5355 /note="assembly_name:Contig35"  
misc_feature 5456. 6757 /note="assembly_name:Contig39"  
misc_feature 6858. 8570 /note="assembly_name:Contig40"  
misc_feature 8671. 9925 /note="assembly_name:Contig41"  
misc_feature 10026. 11426 /note="assembly_name:Contig42"  
misc_feature 11527. 13266 /note="assembly_name:Contig44"  
misc_feature 13367. 14794 /note="assembly_name:Contig47"  
misc_feature 14895. 16054 /note="assembly_name:Contig48"  
misc_feature 16155. 17395 /note="assembly_name:Contig49"  
misc_feature 17496. 19287 /note="assembly_name:Contig50"  
misc_feature 19388. 21294 /note="assembly_name:Contig51"  
misc_feature 21395. 22944 /note="assembly_name:Contig52"  
misc_feature 23045. 24421 /note="assembly_name:Contig53"  
misc_feature 24522. 25870 /note="assembly_name:Contig54"  
misc_feature 25971. 27230 /note="assembly_name:Contig55"  
misc_feature 27331. 28778
```

```
misc_feature /note="assembly_name:Contig57"  
misc_feature 28879. 30893 /note="assembly_name:Contig58"  
misc_feature 30994. 32460 /note="assembly_name:Contig59"  
misc_feature 32561. 33984 /note="assembly_name:Contig60"  
misc_feature 34085. 35285 /note="assembly_name:Contig61"  
misc_feature 35386. 37184 /note="assembly_name:Contig62"  
misc_feature 37285. 39172 /note="assembly_name:Contig63"  
misc_feature 39273. 40874 /note="assembly_name:Contig64"  
misc_feature 40975. 42893 /note="assembly_name:Contig65"  
misc_feature 42994. 44384 /note="assembly_name:Contig66"  
misc_feature 44485. 45999 /note="assembly_name:Contig67"  
misc_feature 46100. 48669 /note="assembly_name:Contig68"  
misc_feature 48770. 50798 /note="assembly_name:Contig69"  
misc_feature 50899. 52809 /note="assembly_name:Contig70"  
misc_feature 52910. 55127 /note="assembly_name:Contig71"  
misc_feature 55228. 58087 /note="assembly_name:Contig72"  
misc_feature 58188. 61004 /note="assembly_name:Contig73"  
misc_feature 61105. 64185 /note="assembly_name:Contig74"  
misc_feature 64286. 67105
```

Query Match 100.0%; Score 20; DB 2; Length 99819;
Best Local Similarity 100.0%; Pred. No. 3;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 gtgaatgcacacttgacat 20
Db 87022 GTGAATGCACACTTTCACAT 87003

RESULT 5
AX119933 11492 bp DNA Linear PAT 11-MAY-2001
LOCUS Sequence 3 from Patent W00129266.
DEFINITION AX119933
ACCESSION AX119933
VERSION AX119933.1 GI:14036679
KEYWORDS

SOURCE
house mouse.
Mus musculus

REFERENCE
1 (bases 1 to 11492)
AUTHORS Hudson,T.J., Engert,J. and Richter,A.
TITLE Identification of arcs mutations and methods of use therefor
JOURNALS Patent: WO 0129266-A 3 26-APR-2001;
MCGILL UNIVERSITY (CA) ; Hopital Sainte-Justine (CA)
LOCATION/Qualifiers

FEATURES

source 1. 11492
/organism="Mus musculus"
/db_xref="taxon:10090"
BASE COUNT 3599 a 2280 c 2387 g 3226 t
ORIGIN

Query Match 92.0%; Score 18.4; DB 6; Length 11492;
Best Local Similarity 95.0%; Pred. No. 22;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gtaaatggccacttgact 20
 ||||||||||||||||
 Db 6396 GTGAATGGCCACTTTCCTCT 6415

RESULT 6
 AF193557 11493 bp DNA linear ROD 07-FEB-2000
 LOCUS Mus musculus sacsin gene, complete cds.
 DEFINITION AF193557.1 GI:6907043
 VERSION
 KEYWORDS
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
 AUTHORS Engert,J.C., Berube,P., Mercier,J., Dore,C., Lepage,P., Ge,B.,
 1 (bases 1 to 11493)
 Bouchard,J.P., Mathieu,J., Melancon,S.B., Schalling,M.,
 Lander,E.S., Morgan,K., Hudson,T.J. and Richter,A.
 ARSACS, a spastic ataxia common in northeastern Quebec, is caused
 by mutations in a new gene encoding an 11.5-Kb ORF
 Nat. Genet. 24 (2), 120-125 (2000)
 20120709
 2 (bases 1 to 11493)
 Engert,J.C., Berube,P., Dore,C., Lepage,P., Ge,B., Hudson,T.J. and
 Richter,A.
 Direct Submission
 Submitted (08-Oct-1999) Genome Centre, Montreal General Hospital,
 1650 Cedar Ave., Montreal, QC H3G 1A4, Canada
FEATURES
 source 1. 11493
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 <1. >11493
 /product="sacsin"
 1. 11493
 /note="molecular chaperone"
 /codon_start=1
 /product="sacsin"
 /protein_id="AAF31263.1"
 /db_xref="GI:6907044"
 /translation="MNTFMPGRELIVVOMYFSEDRHRHPSLWLMKWNKLYTHFSEDL
 TLFDEMPILPTLINDOTCVLIRLIPSVYILIDEPFAOLPEPLADIVOKGLGVL
 KRLDSTIOPPLVKKYIHSPLEPAIIQIMKRIPLQKCNDAISLITPHDAKPLASL
 TDISEKRTIOELTIFKRIHNSDQISSYTKLGGCVLDHFAKLPDRLRSVSD
 SSDEATIRLANMLIEKLTQSCLEFLVDIGNAEYTOEVLQMLMILENLSLKE
 NSNVLDWLPKLFHMSOGHVAAGDLDPDIEVLDLFEVNEEACFPPTITSPDIL
 HSLRQIGLNESLSKERDVYVARKIEALVQSSONODVLMKAKTLLLVKNQTL
 OSSEGMALKKIKWPACKERPNNPGSLVMGDCINCAAPPDMCDAHVLVGSSTP
 LVESYVNLLEQALSTFKETIYAVLKHRTTYVDWTSKTFSEDTYQRPHTILEYGF
 MHDILSEGDSTKALFPMWTKNFCPLAQAVIKPTDLDLQPLVLYNPKTMARFQ
 LEKAGSIEELSDHSIOWYIKVYLSDELSEESKQMLMLIMRLVYNQIPAS
 PNTVPVYIHSRNPSTKLVMPPIHECCYCDIKVDLNDLEDSEVPIILVEDIPMTAE
 WLKVPCLSTRLINPENMGFEQSGREPLTVRKNTLIEEVSDFELKELQVADANA
 TSCSMIMRRMMDIRENLDPMAGHCKPNSNPFSSDPLNTRIKGKISLRG
 EYDKYKGCIGCNVYHTDPIIMASREPMEDPNINISHIKIDRSRGKIKMSK
 OOKRLRFPNOKRPIDVFCOLPLAVEAPVSYNGFLRSLSTRQDEAVSEVSTCY
 NTADITYSLDEFSLGHRILITQSVNSVYLYKLIETNPISLADDTIILKKRVCSK
 ALNAVLYSLDEAKALMKTCSSNNKLPDVKVSCIIQITVEEHHVFRADLQSP
 LERGGDDPATLEFMAKSGOSKPSDELPOKVDCTWMLICMDTEGALKSLNMSG
 BRGLVPGGAVGLVLEHTQEOEKWVYKPHIGEVCFPLRIKIGLPIHNGCPAVNSR
 KEIKMTDTRKGRNNTFMRIYKATLQALSVALDLAGELDTDYTYANWPPDLVHD
 DSVYICGTEYDIAGKGLKELTVFSDGSMVSMKNVRLLDSIIQKRVGSAARFIE
 LKYLKTSKNLCALVPSYKAGEEACQOILENTEFSEQFSEVFPNIOETEA
 ELRDLPMNVLEKLEDEFGILRTPVCPVCSLEGHPLVPSRLIPEGVARLPTKD
 GREPGYODVYLNELIILKVLQAGKADILMDMLERASVAEINKSDHAACLRSS
 ILLSLIDEKLTIKDPRANDFAKQOTIPLRPLTPAGSLKMKGSPFEMPATD
 IYTAEDIVLCLOPLINLNSFRGCGSVSLAVKEFGLLKKFPYDVLINQKQVAK
 SYDDGSTITQENTNACRYLHEAVLQNMARATIIIEKIPCFYLLIVEVYSESVS
 FLNFEAAVYLVQLPNPKYKNNRELFEVSQVSTVEDFVALVLSIDDERKQJOTE
 ENFOLCRRIISGIIWLSLIREKQEFCEKNYKILLPDTNLLLPKSLCYNDCPMIKY

BASE COUNT 3599 a 2281 c 2387 g 3226 t
 ORIGIN
 Query Match 92.0%; Score 18.4; DB 10; Length 11493;
 Best Local Similarity 95.0%; Pred. No. 22;
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gtaaatggccacttgact 20
 ||||||||||||||||
 Db 6397 GTGAATGGCCACTTTCCTCT 6416

RESULT 7
 G37224 434 bp DNA linear STS 31-MAR-1998
 LOCUS SHGC-57260 Human Homo sapiens STS genomic, sequence tagged site.
 DEFINITION G37224
 ACCESSION G37224.1 GI:2996875
 VERSION
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
 AUTHORS Myers,R.M.
 1 (bases 1 to 434)
TITLE Human STS (1997)
JOURNAL Unpublished
COMMENT
 Contact: Richard M. Myers
 Stanford Human Genome Center (SHGC)
 Stanford University School of Medicine
 Department of Genetics, M-344, Stanford, CA 94305, USA
 Tel: 4157259687
 Fax: 4157259689
 Email: myers@shgc.stanford.edu

Primer A: TTTTTCAGGCGCTGTGTC
Primer B: CATGTTCTCCAGATGCCA
STS size: 99
PCR Profile:

Initial incubation: 95 degrees C for 10 minutes

Denaturation: 94 degrees C for 30 seconds
Annealing: 60 degrees C for 30 seconds
Polymerization: 72 degrees C for 23 seconds

PCR Cycles: 30
Thermal Cycler: Perkin Elmer 9700

Protocol:

Template: 25 ng
Primer: each 1 uM
dNTPs: each 200 uM
AmpliTaq Gold Polymerase: 0.07 units/uL
Total Vol: 5 uL

Buffer:

MgCl2: 2.5 mM
KCl: 50 mM
Tris-HCl: 10 mM
pH: 8.3

Prepared with primer pairs derived from W60474 -- Unigene.

FEATURES
source location/Qualifiers

1..434
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="10"
/clone_1b="Human"

STS
primer_bind 3..22

primer_bind complement(82..101)

BASE COUNT 98 a 93 c 97 g 146 t

ORIGIN

Query Match

Best Local Similarity 87.0%; Score 17.4; DB 11; Length 434;
Pred. No. 69;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 tgaatggccactgtgac 20

||||| ||||||| |||||

Db 301 TGAATTGCCACTTGCAC 319

RESULT 8

AF113358

LOCUS AF113358 766 bp DNA linear INV 27-APR-2000

DEFINITION Ips latidens haplotype 1 cytochrome oxidase I gene, partial cds;

AF113358 mitochondrial gene for mitochondrial product.

AF113358.1 GI:7650044

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

2 (bases 1 to 766)

REFERENCE

AUTHORS

TITLE

JOURNAL

Submitted (15-DEC-1998) ESPM-Division of Insect Biology, University

of California, Berkeley, 201 Wellman Hall, Berkeley, CA 94720, USA

Location/Qualifiers

source

1..766
/organism="Ips latidens"
/organism="mitochondrion"
/db_xref="taxon:102838"
/haplotype="1"

CDS

/note="similar to Drosophila yakuba cytochrome oxidase I"
/codon_start=2
/transl_table=5
/product="cytochrome oxidase I"
/protein_id="AAF65880.1"
/db_xref="GI:7650045"

/translation="PQGLISHIGGSGKKEAFVGMIVYNTATGLGFFVMAHHM
FTIGMDVDTRAYTSATMTIATPTGKITFSWATPFGTQISFSSISLGFLEFTM
GGTGTGVLNNSIDITLHDPTVYVAHFHVLVSMGAFATLACIYQWPLFTGLTKK
YKTHPLTMEVGNLTFFPQHFGLSGMFRRSYDPAVLLMNTISSISMSLSIVF
YFIFLMEFSAGRSKISALNLSLSLEWLYLPPSDH"

BASE COUNT 218 a 156 c 122 g 270 t

ORIGIN

Query Match

Best Local Similarity 87.0%; Score 17.4; DB 3; Length 766;
Pred. No. 71;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 gtaatggccactgtgac 19

||||| ||||||| |||||

Db 223 GTAATGCCACTTTTCAC 241

RESULT 9

AF113359

LOCUS AF113359 766 bp DNA linear INV 27-APR-2000

DEFINITION Ips latidens haplotype 2 cytochrome oxidase I gene, partial cds;

AF113359 mitochondrial gene for mitochondrial product.

AF113359.1 GI:7650046

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

2 (bases 1 to 766)

REFERENCE

AUTHORS

TITLE

JOURNAL

Submitted (15-DEC-1998) ESPM-Division of Insect Biology, University

of California, Berkeley, 201 Wellman Hall, Berkeley, CA 94720, USA

Location/Qualifiers

source

CDS

/note="similar to Drosophila yakuba cytochrome oxidase I"
/codon_start=2
/transl_table=5
/product="cytochrome oxidase I"
/protein_id="AAF65881.1"
/db_xref="GI:7650047"

/translation="PQGLISHIGGSGKKEAFVGMIVYNTATGLGFFVMAHHM
FTIGMDVDTRAYTSATMTIATPTGKITFSWATPFGTQISFSSISLGFLEFTM
GGTGTGVLNNSIDITLHDPTVYVAHFHVLVSMGAFATLACIYQWPLFTGLTKK
YKTHPLTMEVGNLTFFPQHFGLSGMFRRSYDPAVLLMNTISSISMSLSIVF
YFIFLMEFSAGRSKISALNLSLSLEWLYLPPSDH"

```

BASE COUNT      220 a      155 c      120 g      271 t
ORIGIN
Query Match      87.0%; Score 17.4; DB 3; Length 766;
Best Local Similarity 94.7%; Pred. No. 71;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY      1      tgaatgccacattgcac 19
        |||
        223 GTGAATGCCACATTGCAC 241

RESULT 10
AC069021      175309 bp      DNA      linear      HTG 30-NOV-2001
LOCUS      Homo sapiens chromosome 10 clone RP11-254K3, WORKING DRAFT
DEFINITION      SEQUENCE, 2 unordered pieces.
ACCESSION      AC069021
VERSION      AC069021.8      GI:17155019
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULFILL; HTGS_ACTIVEFIN.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE      1 (bases 1 to 175309)
AUTHORS      Smith,D.R.
TITLE      Genome Therapeutics Corporation Sequencing Center: Human Genome
JOURNAL      Unpublished
AUTHORS      2 (bases 1 to 175309)
TITLE      Smith,D.R.
COMMENT      Submitted (17-MAY-2000) Genome Therapeutics Corporation, 100 Beaver
              Street, Waltham, MA 02453, USA
              On Nov 30, 2001 this sequence version replaced gi:14277222.
              -----
              Genome Center
              Center: Genome Therapeutics Corporation
              Center code: GTC
              Web site: http://www.genomecorp.com/
              Contact: gtc-seqcenter@genomecorp.com
              -----
              Project Information
              Center project name: hg364
              -----
              Summary Statistics
              Sequencing vector: N/A
              Chemistry: Dye-terminator Big Dye; 100% of reads
              Assembly program: Phrap; version 990315
              Consensus quality: 174470 bases at least Q40
              Consensus quality: 174909 bases at least Q30
              Consensus quality: 175102 bases at least Q20
              Insert size: 175308; sum-of-contrigs
              Quality coverage: 7.4x in Q20 bases; sum-of-contrigs
              -----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contrigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contrigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1      9697: contrig of 9697 bp in length
* 9698      9797: gap of unknown length
* 9798      175309: contrig of 165512 bp in length.
FEATURES
Source
1..175309
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-254K3"
/clone_lib="RPCT-11"
1..9697
/note="assembly_name:Contrig1
misc_feature

```

```

misc_feature      clone_end:SP6"
9798..175309
/note="assembly_name:Contrig2
clone_end:77"
BASE COUNT      54898 a      36843 c      34823 g      48645 t      100 others
ORIGIN
Query Match      87.0%; Score 17.4; DB 2; Length 175309;
Best Local Similarity 94.7%; Pred. No. 91;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY      2      tgaatgccacattgcac 20
        |||
        Db 14083 TGAATGCCACATTGCAC 14101

RESULT 11
AP000751      176713 bp      DNA      linear      PRI 22-MAR-2001
LOCUS      Homo sapiens genomic DNA, chromosome 11q, clone:RP11-679A11,
DEFINITION      complete sequence.
ACCESSION      AP000751
VERSION      AP000751.4      GI:13429919
KEYWORDS      HTG.
SOURCE      Homo sapiens DNA, clone:RP11-679A11.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE      1 (sites)
AUTHORS      Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
              Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
              Homo sapiens genomic DNA
              Published only in Database (1999) In press
              2 (bases 1 to 176713)
              Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
              Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
              Direct Submission
              Submitted (25-NOV-1999) Masahira Hattori, The Institute of Physical
              and Chemical Research (RIKEN), Genomic Sciences Center (GSC),
              1-7-22 Suenho-cho, Tsukuba, Ibaraki, Japan, 305-8565, Japan
              (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
              Tel:81-45-503-9111, Fax:81-45-503-9170)
              On Mar 21, 2001 this sequence version replaced gi:9757498.
              -----
              Location/Qualifiers
              1..176713
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /chromosome="11"
              /map="11q"
              /clone="RP11-679A11"
BASE COUNT      58670 a      33938 c      32613 g      51492 t
ORIGIN
Query Match      87.0%; Score 17.4; DB 9; Length 176713;
Best Local Similarity 94.7%; Pred. No. 91;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY      2      tgaatgccacattgcac 20
        |||
        Db 68285 TGAATGCCACATTGCAC 68303

RESULT 12
AP001265      192913 bp      DNA      linear      HTG 30-MAY-2000
LOCUS      Homo sapiens chromosome 11 clone RP11-741L23 map 11q24, WORKING
DEFINITION      DRAFT SEQUENCE, 34 unordered pieces.
ACCESSION      AP001265
VERSION      AP001265.2      GI:8117657
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens DNA, clone:RP11-741L23.

```

ORGANISM Homo sapiens
Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
Mammalia: Eutheria: Primates: Catarrhini: Hominiidae: Homo.
REFERENCE 1 (bases 1 to 192913)
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Homo sapiens 192,913 genomic DNA of 11q24
JOURNAL Published Only in Database (2000) In press
AUTHORS 2 (bases 1 to 192913)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Submitted (23-FEB-2000) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
Japan (E-mail:hattori@gsc.riken.go.jp,
URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
Fax:81-42-778-9924)
COMMENT On May 31, 2000 this sequence version replaced gi:7106133.
----- Genome Center
Center: RIKEN Genomic Sciences Center(GSC)
Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/
Contact: hattori@gsc.riken.go.jp
----- Project Information
Center project name: Humdrift1
Center clone name: RP11-741L23
----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 170147 bases at least Q40
Consensus quality: 181282 bases at least Q30
Consensus quality: 186868 bases at least Q20
Insert size: 189613; sum-of-ctrls
Quality coverage: 4.30x in Q20 bases; sum-of-ctrls

NOTE: This is a 'working draft' sequence. It currently consists of
34 contigs. The true order of the pieces is not known and their
order in this sequence record is arbitrary. Gaps between the
contigs are represented as runs N, but the exact sizes of the gaps
are unknown. This record will be updated with the finished sequence
as soon as it is available and the accession number will be
preserved
1 15621 contig of 15621 bp in length
15722 30722 contig of 15001 bp in length
30823 45308 contig of 14486 bp in length
45409 56803 contig of 11395 bp in length
56904 64499 contig of 7596 bp in length
64600 74039 contig of 9440 bp in length
74140 83223 contig of 9090 bp in length
83330 92098 contig of 8769 bp in length
92199 99531 contig of 7333 bp in length
99632 108069 contig of 8438 bp in length
108170 113991 contig of 5721 bp in length
113991 120625 contig of 6635 bp in length
120726 126591 contig of 5866 bp in length
126692 133471 contig of 6780 bp in length
133471 139936 contig of 6365 bp in length
139936 140036 contig of 100 bp in length
140036 145226 contig of 5190 bp in length
145226 145326 contig of 100 bp in length
145326 148567 contig of 3241 bp in length
148568 148667 contig of 100 bp in length
148668 152872 contig of 4205 bp in length
152873 152972 contig of 100 bp in length
152973 156332 contig of 3360 bp in length
156333 156432 contig of 100 bp in length
156433 160140 contig of 3708 bp in length
160141 160240 contig of 100 bp in length
160241 163773 contig of 3533 bp in length
163774 163873 contig of 100 bp in length
163874 168073 contig of 4200 bp in length
168074 168173 contig of 100 bp in length
168174 172134 contig of 3961 bp in length
172135 172234 contig of 100 bp in length
172235 174914 contig of 2680 bp in length
174915 175014 contig of 100 bp in length
175015 177403 contig of 2389 bp in length
177404 177503 contig of 100 bp in length
177504 178982 contig of 1479 bp in length
178983 179082 contig of 100 bp in length
179084 181480 contig of 2398 bp in length
181481 181580 contig of 100 bp in length
181581 182602 contig of 1022 bp in length
182603 182702 contig of 100 bp in length
182703 184867 contig of 2165 bp in length
184868 184967 contig of 100 bp in length
184968 186523 contig of 1556 bp in length

184968 186523 contig of 1556 bp in length
186524 187814 contig of 1191 bp in length
187915 190302 contig of 2388 bp in length
190403 191751 contig of 1349 bp in length
191852 192913 contig of 1062 bp in length
Sequence updated (26-May-2000).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 34 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 15621: contig of 15621 bp in length
* 15622 15721: gap of 100 bp
* 15722 30722: contig of 15001 bp in length
* 30723 30822: gap of 100 bp
* 30823 45308: contig of 14486 bp in length
* 45309 45408: gap of 100 bp
* 45409 56803: contig of 11395 bp in length
* 56804 56903: gap of 100 bp
* 56904 64499: contig of 7596 bp in length
* 64500 64599: gap of 100 bp
* 64600 74039: contig of 9440 bp in length
* 74040 74139: gap of 100 bp
* 74140 83223: contig of 9090 bp in length
* 83230 83329: gap of 100 bp
* 83330 92098: contig of 8769 bp in length
* 92099 92198: gap of 100 bp
* 92199 99531: contig of 7333 bp in length
* 99532 99631: gap of 100 bp
* 99632 108069: contig of 8438 bp in length
* 108070 108169: gap of 100 bp
* 108170 113890: contig of 5721 bp in length
* 113891 113990: gap of 100 bp
* 113991 120625: contig of 6635 bp in length
* 120626 120725: gap of 100 bp
* 120726 126591: contig of 5866 bp in length
* 126592 126691: gap of 100 bp
* 126692 133471: contig of 6780 bp in length
* 133472 133571: gap of 100 bp
* 133572 139936: contig of 6365 bp in length
* 139937 140036: gap of 100 bp
* 140037 145226: contig of 5190 bp in length
* 145227 145326: gap of 100 bp
* 145327 148567: contig of 3241 bp in length
* 148568 148667: gap of 100 bp
* 148668 152872: contig of 4205 bp in length
* 152873 152972: gap of 100 bp
* 152973 156332: contig of 3360 bp in length
* 156333 156432: gap of 100 bp
* 156433 160140: contig of 3708 bp in length
* 160141 160240: gap of 100 bp
* 160241 163773: contig of 3533 bp in length
* 163774 163873: gap of 100 bp
* 163874 168073: contig of 4200 bp in length
* 168074 168173: gap of 100 bp
* 168174 172134: contig of 3961 bp in length
* 172135 172234: gap of 100 bp
* 172235 174914: contig of 2680 bp in length
* 174915 175014: gap of 100 bp
* 175015 177403: contig of 2389 bp in length
* 177404 177503: gap of 100 bp
* 177504 178982: contig of 1479 bp in length
* 178983 179082: gap of 100 bp
* 179084 181480: contig of 2398 bp in length
* 181481 181580: gap of 100 bp
* 181581 182602: contig of 1022 bp in length
* 182603 182702: gap of 100 bp
* 182703 184867: contig of 2165 bp in length
* 184868 184967: gap of 100 bp
* 184968 186523: contig of 1556 bp in length

FEATURES
Source
1.192913
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q24"
/clone="RP11-741L23"
1.15621
misc_feature
/note="assembly-fragment"
15722..30722
/note="assembly-fragment"
30823..45308
/note="assembly-fragment"
45409..56803
/note="assembly-fragment"
56904..64499
/note="assembly-fragment"
64600..74039
/note="assembly-fragment"
74140..83229
/note="assembly-fragment"
83330..92098
/note="assembly-fragment"
92199..99531
/note="assembly-fragment"
99632..108069
/note="assembly-fragment"
108170..113890
/note="assembly-fragment"
113991..120625
/note="assembly-fragment"
120726..126591
/note="assembly-fragment clone_end:T7 vector_side:left"
126592..133471
/note="assembly-fragment"
133572..139936
/note="assembly-fragment"
140037..145226
/note="assembly-fragment"
misc_feature
Query Match 87.0%; Score 17.4; DB 2; Length 192913;
Best Local Similarity 94.7%; Pred. No. 92;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2 tgaatggccacttgcact 20
|||||
Db 68804 TGAATGGCCACTTGCATT 68786

RESULT 13
AC073587/ C 216521 bp DNA linear PRI 28-JUL-2001
DEFINITION Homo sapiens chromosome 10 clone RP11-572P18, complete sequence.
ACCESSION AC073587
VERSION AC073587.5 GI:15027737
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 216521)
AUTHORS Smith, D.R.
TITLE Genome Therapeutics Corporation Sequencing Center: Human Genome
JOURNAL . Unpublished

REFERENCE 2 (bases 1 to 216521)
AUTHORS Smith, D.R.
TITLE Direct Submission
JOURNAL Submitted (25-JUN-2000) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02453, USA
REFERENCE 3 (bases 1 to 216521)
AUTHORS Smith, D.R.
TITLE Direct Submission
JOURNAL Submitted (28-JUL-2001) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02453, USA
COMMENT On Jul 28, 2001 this sequence version replaced gi:14787184.
FEATURES
Source
1.216521
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-572P18"
/clone_lib="RPCT-11"
BASE COUNT 61858 a 44519 c 46794 g 63350 t
ORIGIN

Query Match 87.0%; Score 17.4; DB 9; Length 216521;
Best Local Similarity 94.7%; Pred. No. 92;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2 tgaatggccacttgcact 20
|||||
Db 54172 TGAATGGCCACTTGCATT 54154

RESULT 14
AC019279 241392 bp DNA linear HTG 23-SEP-2000
LOCUS AC019279
DEFINITION Homo sapiens clone RP11-14E3, *** SEQUENCING IN PROGRESS ***; 74
unordered pieces.
ACCESSION AC019279
VERSION AC019279.5 GI:10280840
KEYWORDS HTG: HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 241392)
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Boguslavskiy, L., Bouckhalter, B., Brown, A., Burkett, G., Castle, A.,
Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,
Dearellano, K., Dewar, K., Domino, M., Doyle, M., Fenestor, J.,
Ferreira, P., Fitzhugh, W., Forrest, C., Gage, D., Galsagan, J.,
Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
Landers, T., Lehocck, J., Levine, R., Lieu, C., Liu, G., Locke, K.,
Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K.,
McPheeters, R., Meldrum, J., Meneus, L., Morrow, J., Naylor, J.,
Norman, C. H., O'Connor, T., O'Donnell, P., Olivar, T. M., Peterson, K.,
Pierre, N., Pisanl, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,
Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,
Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J.,
Zimmer, A., and Zody, M.
DIRECT SUBMISSION
TITLE Direct Submission
JOURNAL Submitted (31-DEC-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Sep 23, 2000 this sequence version replaced gi:7630670.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RW/RepeatMasker.html
----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information -----
Center project name: L3393
Center Clone name: 14_E_3

* NOTE: This is a 'working draft' sequence. It currently
* consists of 74 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1
* 4479 4578: contig of 4478 bp in length
* 4479 4578: gap of 100 bp
* 4579 6006: contig of 1428 bp in length
* 6007 6106: gap of 100 bp
* 6107 7245: contig of 1139 bp in length
* 7246 7345: gap of 100 bp
* 7346 8403: contig of 1058 bp in length
* 8404 8503: gap of 100 bp
* 8504 9605: contig of 1102 bp in length
* 9606 9705: gap of 100 bp
* 9706 10965: contig of 1260 bp in length
* 10966 11065: gap of 100 bp
* 11066 12108: contig of 1043 bp in length
* 12109 12208: gap of 100 bp
* 12209 13446: contig of 1238 bp in length
* 13447 13546: gap of 100 bp
* 13547 14879: contig of 1333 bp in length
* 14880 14979: gap of 100 bp
* 14980 16230: contig of 1251 bp in length
* 16231 16330: gap of 100 bp
* 16331 17535: contig of 1205 bp in length
* 17536 17635: gap of 100 bp
* 17636 18689: contig of 1054 bp in length
* 18690 18789: gap of 100 bp
* 18790 20178: contig of 1389 bp in length
* 20179 20278: gap of 100 bp
* 20279 21831: contig of 1553 bp in length
* 21832 21931: gap of 100 bp
* 21932 23140: contig of 1209 bp in length
* 23141 23240: gap of 100 bp
* 23241 24600: contig of 1360 bp in length
* 24601 24700: gap of 100 bp
* 24701 26439: contig of 1739 bp in length
* 26440 26539: gap of 100 bp
* 26540 27663: contig of 1124 bp in length
* 27664 27763: gap of 100 bp
* 27764 29674: contig of 1911 bp in length
* 29675 29774: gap of 100 bp
* 29775 31060: contig of 1286 bp in length
* 31061 31160: gap of 100 bp
* 31161 32996: contig of 1836 bp in length
* 32997 33096: gap of 100 bp
* 33097 34353: contig of 1257 bp in length
* 34354 34453: gap of 100 bp
* 34454 36377: contig of 1924 bp in length
* 36378 36477: gap of 100 bp
* 36478 38540: contig of 2063 bp in length
* 38541 38640: gap of 100 bp
* 38641 40445: contig of 1805 bp in length
* 40446 40545: gap of 100 bp
* 40546 51430: contig of 10885 bp in length
* 51431 51530: gap of 100 bp
* 51531 53722: contig of 2192 bp in length
* 53723 53822: gap of 100 bp
* 53823 55996: contig of 2174 bp in length
* 55997 56096: gap of 100 bp
* 56097 57762: contig of 1666 bp in length

* 57763 57862: gap of 100 bp
* 57863 59701: contig of 1839 bp in length
* 59702 59801: gap of 100 bp
* 59802 61877: contig of 2076 bp in length
* 61878 61977: gap of 100 bp
* 61978 63529: contig of 1552 bp in length
* 63530 63629: gap of 100 bp
* 63630 65374: contig of 1745 bp in length
* 65375 65474: gap of 100 bp
* 65475 67237: contig of 1763 bp in length
* 67238 67337: gap of 100 bp
* 67338 68940: contig of 1603 bp in length
* 68941 69040: gap of 100 bp
* 69041 71146: contig of 2106 bp in length
* 71147 71246: gap of 100 bp
* 71247 73303: contig of 2057 bp in length
* 73304 73403: gap of 100 bp
* 73404 75790: contig of 2387 bp in length
* 75791 75890: gap of 100 bp
* 75891 78184: contig of 2294 bp in length
* 78185 78284: gap of 100 bp
* 78285 81055: contig of 2771 bp in length
* 81056 81155: gap of 100 bp
* 81156 83821: contig of 2666 bp in length
* 83822 83921: gap of 100 bp
* 83922 85940: contig of 2019 bp in length
* 85941 86040: gap of 100 bp
* 86041 88683: contig of 2643 bp in length
* 88684 88783: gap of 100 bp
* 88784 91205: contig of 2422 bp in length
* 91206 91305: gap of 100 bp
* 91306 94475: contig of 3170 bp in length
* 94476 94575: gap of 100 bp
* 94576 965917: contig of 2242 bp in length
* 96592 97017: gap of 100 bp
* 97018 99382: contig of 2365 bp in length
* 99383 99482: gap of 100 bp
* 99483 102002: contig of 2520 bp in length
* 102003 102102: gap of 100 bp
* 102103 105794: contig of 3692 bp in length
* 105795 105894: gap of 100 bp
* 105895 108858: contig of 2964 bp in length
* 108859 108958: gap of 100 bp
* 108959 113547: contig of 4589 bp in length
* 113548 113647: gap of 100 bp
* 113648 119743: contig of 6096 bp in length
* 119744 119843: gap of 100 bp
* 119844 122714: contig of 2871 bp in length
* 122715 122814: gap of 100 bp
* 122815 127700: contig of 4886 bp in length
* 127701 127800: gap of 100 bp
* 127801 131089: contig of 3289 bp in length
* 131090 131189: gap of 100 bp
* 131190 135083: contig of 3894 bp in length
* 135084 135183: gap of 100 bp
* 135184 139232: contig of 4049 bp in length
* 139233 139332: gap of 100 bp
* 139333 143122: contig of 3790 bp in length
* 143123 143222: gap of 100 bp
* 143223 148113: contig of 4891 bp in length
* 148114 148213: gap of 100 bp
* 148214 153325: contig of 5112 bp in length
* 153326 153425: gap of 100 bp
* 153426 158584: contig of 5159 bp in length
* 158585 158684: gap of 100 bp
* 158685 163480: contig of 4796 bp in length
* 163481 163580: gap of 100 bp
* 163581 168903: contig of 5323 bp in length
* 168904 169003: gap of 100 bp
* 169004 175183: contig of 6180 bp in length
* 175184 175283: gap of 100 bp
* 175284 180173: contig of 4890 bp in length
* 180174 180273: gap of 100 bp

* 180274 184306: contig of 4033 bp in length
* 184307 184406: gap of 100 bp
* 184407 189795: contig of 5389 bp in length
* 189796 189895: gap of 100 bp
* 189896 198138: contig of 8243 bp in length
* 198139 198238: gap of 100 bp
* 198239 206826: contig of 8588 bp in length
* 206827 206926: gap of 100 bp
* 206927 214548: contig of 7622 bp in length
* 214549 214648: gap of 100 bp
* 214649 221820: contig of 7172 bp in length
* 221821 221920: gap of 100 bp
* 221921 230400: contig of 8480 bp in length
* 230401 230500: gap of 100 bp

Query Match 87.0%; Score 17.4; DB 2; Length 241392;
Best Local Similarity 94.7%; Pred. No. 93;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 tgaatggccacttgcact 20
Db 79271 TGAATGGCCACTTTGCATT 79253
|||||

RESULT 15
LOCUS POU95073 993 bp DNA linear INV 02-FEB-2000
DEFINITION Pisaster ochraceus cytochrome oxidase I (COI) gene, partial cds;
ACCESSION U95073
VERSION U95073.2 GI:6855152
KEYWORDS
SOURCE ocreous starfish.
ORGANISM Mitochondrion Pisaster ochraceus
Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Asterozoa;
Asteroidea; Forcipulatacea; Forcipulata; Asteriidae; Pisaster.

REFERENCE
AUTHORS Hirnecvich,A.W., Rocha-Olivares,A. and Foltz,D.W.
TITLE 1 (bases 1 to 993)
Phylogenetic analysis of molecular lineages in a speciose subgenus
of sea stars (Leptasterias subgenus Hexasterias)

JOURNAL
REFERENCE Unpublished
AUTHORS 2 (bases 1 to 993)
TITLE Hirnecvich,A.W. and Foltz,D.W.
JOURNAL Direct Submission
Submitted (21-MAR-1997) Zoology and Physiology, Louisiana State
University, South Campus Drive, Baton Rouge, LA 70803-1725, USA
3 (bases 1 to 993)
REFERENCE Hirnecvich,A.W., Rocha-Olivares,A. and Foltz,D.W.
TITLE Direct Submission
Submitted (02-FEB-2000) Zoology and Physiology, Louisiana State
University, South Campus Drive, Baton Rouge, LA 70803-1725, USA

REMARK
COMMENT Sequence update by submitter
FEATURES
On Feb 2, 2000 this sequence version replaced gi:2105417.
Location/Qualifiers
1..993

gene /organism="Pisaster ochraceus"
CDS /organeller="mitochondrion"
/specimen_voucher="#I-302, Louisiana State University"
/db_xref="taxon:7612"
<1..>993
/gene="COI"
/gene="COI"
<1..>993
/codon_start=1
/transl_table=9
/product="cytochrome oxidase I"
/protein_id="AA58051.2"
/db_xref="GI:6855153"

/translation="GNWLIPLMIGAPDMAFPMNNMFWLIPSPILLASAGVESGA
GTGWTIPPLSSGLAHAGSVLAIFSLACASSLASLAINFTITNNRTPGMSFDR
LPLFWSVFVTLFLDLPLVLAGAITMLLTDRNINITEFDPAAGGDPILFQHLWEF
GHPNVYILLPFGMIASHVIAHYAGNPEFGYALISIGILGFLWAHMFVIG
MDVNTRAYFTATMTIAVPTGIKVENMATLOGNNLRMDTPLLALGFEVLEFTIGILT

BASE COUNT 255 a 225 c 184 g 329 t
ORIGIN GVVLANSSIDIILHETYYVVAHFHYVLSMGAVFAIFAGFTHMFPLESSVSLHPLW"

Query Match 85.0%; Score 17; DB 3; Length 993;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 tgaatggccacttgcac 18
Db 733 TGAATGGCCACTTTGCA 749
|||||

Search completed: May 22, 2002, 06:59:46
Job time: 9171 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 22, 2002, 08:46:53 ; Search time 3328.52 Seconds
(without alignments)
3728.214 Million cell updates/sec

Title: US-09-693-205-7_COPY_12200_12792

Perfect score: 593
Sequence: 1 acatctatgtttacaggct.....accctcacatttatgtctt 593

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1797656 seqs, 10463268293 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-Processing: Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_hcgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result Query
No. Score Match Length DB ID Description

1	593	100.0	4318	9	AB018273	AB018273 Homo sapi
2	593	100.0	12793	6	AX119931	AX119931 Sequence
3	593	100.0	12793	9	AF193556	AF193556 Homo sapi
4	593	100.0	92693	9	AL157766	AL157766 Human DNA
5	593	100.0	99819	2	AC079761	AC079761 Homo sapi
6	155	26.1	418	11	G36555	G36555 SHGC-53325
7	96	16.2	3289	9	AB056815	AB056815 Macaca fa
8	51	8.6	51	6	AX160575	AX160575 Sequence
9	51	8.6	51	6	AX160577	AX160577 Sequence
10	34	5.7	50	6	AX160578	AX160578 Sequence
11	25	4.2	51	6	AX160576	AX160576 Sequence
12	24	4.0	50	6	AX160574	AX160574 Sequence
13	24	4.0	51	6	AX160573	AX160573 Sequence
14	23	3.9	149128	2	AL670276	AL670276 Mus muscu
15	23	3.9	336028	2	AC092096	AC092096 Mus muscu
16	22	3.7	349980	6	AX344559	AX344559 Sequence
17	21	3.5	1557	4	AF018634	AF018634 Canis fam
18	21	3.5	4604	6	AX278036	AX278036 Sequence
19	21	3.5	4604	6	AX333807	AX333807 Sequence
20	21	3.5	4604	6	AX344874	AX344874 Sequence
21	21	3.5	35920	3	CEC4484	Z79598 Caenorhabdl
22	21	3.5	38418	10	M0SPRNP4	U29186 Mus musculu
23	21	3.5	39760	10	U29187	U29187 Mus musculu
24	21	3.5	70892	2	AC027533	AC027533 Homo sapi
25	21	3.5	71456	10	MMA298054	AL356316 Mus muscu
26	21	3.5	74102	9	AL356316	AL356316 Human DNA
27	21	3.5	79946	2	AL359874	AL359874 Human DNA
28	21	3.5	110000	2	PFMAL13P2_1	Continuation (2 of
29	21	3.5	141794	2	AC093455	AC093455 Homo sapi
30	21	3.5	149901	2	AC022243	AC022243 Homo sapi
31	21	3.5	168957	2	AC025079	AC025079 Homo sapi
32	21	3.5	171279	2	AC036185	AC036185 Homo sapi
33	21	3.5	171721	9	AC022821	AC022821 Homo sapi
34	21	3.5	172639	30	AC009566	AC009566 Homo sapi
35	21	3.5	176333	2	AC108022	AC108022 Homo sapi
36	21	3.5	176863	2	AL358353	AL358353 Homo sapi
37	21	3.5	178100	9	AC009080	AC009080 Homo sapi
38	21	3.5	180526	9	AC027333	AC027333 Homo sapi
39	21	3.5	182725	2	AC021206	AC021206 Homo sapi
40	21	3.5	184529	9	AC009156	AC009156 Homo sapi
41	21	3.5	187047	9	AC018861	AC018861 Homo sapi
42	21	3.5	197551	2	AL358073	AL358073 Homo sapi
43	21	3.5	206187	2	AC069139	AC069139 Homo sapi
44	21	3.5	245997	2	AC093360	AC093360 Mus muscu
45	20	3.4	1947	9	AK023911	AK023911 Homo sapi

ALIGNMENTS

RESULT	1	4318 bp	mrna	linear	PRI 16-JUN-1999
LOCUS	AB018273				
DEFINITION	Homo sapiens mRNA for KIAA0730 protein, partial cds.				
ACCESSION	AB018273				
VERSION	AB018273.1	GI:3882180			
KEYWORDS					
SOURCE					
ORGANISM	Homo sapiens adult male brain cDNA to mRNA, clone_11b:pbluescriptII SK plus clone:hk03632.				
REFERENCE					
AUTHORS	1 (sites)				
TITLE	Nagase,T., Ishikawa,K., Suyama,M., Kikuno,R., Miyajima,N., Tanaka,A., Kotani,H., Nomura,N. and Ohara,O.				
JOURNAL	Prediction of the coding sequences of unidentified human genes. XI. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro				
MEDLINE	DNA Res. 5 (5), 277-286 (1998)				
REFERENCE	99087487				
AUTHORS	2 (bases 1 to 4318)				
TITLE	Ohara,O., Suyama,M., Nagase,T., Ishikawa,K. and Kikuno,R.				
	Direct Submission				

JOURNAL

Submitted (08-Oct-1998) Osamu Ohara, Kazusa DNA Research Institute,
Laboratory of DNA Technology; Yana 1532-3, Kisarazu, Chiba
292-0812, Japan (E-mail:cdhainfo@kazusa.or.jp, Tel:+81-438-52-3913,
Fax:+81-438-52-3914)

FEATURES

source Location/Qualifiers

1..4318
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="h03632"
/sex="male"
/tissue_type="brain"
/clone_lib="pbluescriptII SK plus"
/dev_stage="adult"
1..3015
/gene="K1AA0730"
/gene="K1AA0730"
<1..3015
/codon_start=1
/product="K1AA0730 protein"
/protein_id="BAA34450.1"
/db_xref="GI:3882181"
/translation="VTFTMSWVEFLRNIGLKAYILSSOOLLOFRAKEISVANTEMNSKE
TLQNTVDILHRIROBRDLISGNFLKELSLIPICPREAPAEPIRPHQVEVNTL
PLKNGAQNPKFKQCDVQLMTSCPLPEKATPLSKEDSGSDLGQEDLEVLN
MLNVLDPDLKVINNCNINICNTITLDEEMVTRAKVLRSIYEFLSAEKREFRQLRG
VAFVVEDGDKLKEPEVYINLEYSDEKPYLYKPLLEGLQHFHKLGTEDIISTK
QYEVLSLIFKNSGKQDLPNEMRYVSGLSFRLNSDYKFSVSDLENVDLALYL
PSQDRLVKSSILVEDDAPHYKSRIOGNVOMLDSOCYKQKHGFPTKLIMLEPO
KLRLPLSLISIEEOLDEFPKVCQALCSQGRLOLLSSQFPTGLIRIKMHEHND
AFLANEKRIKRCALRGKAYSCFEKQCTTIRVGVFNPIPHSRETRAFLEKRGNAV
ILYI0HSDSKDINFLALAMTLKSAIDNLISDYSYILMLGCNDIYRIGELDYSYG
KYDSESPKLELPMGPTPIPAEIHYYTLMDPNVYVPEYVGYLVDAEGDIDYGYOP
TYYVAIIVQEVREDADNSFLGKITQIDIGYSEKIVASLDLYKFSREESGQSDS
ASTPTSPTEFLTPGLRSLPLSGRESHKTSKQSPKLVNLSLPELKEVTSVE
QAMKLPSESRKKIIRRLYLKMHDPKNREHDSKQSPKLVNLSLPELKEVTSVE
BASRSTESASRQSDKYSRQREYTSNNOETSIKSEKQOQNKCPSPAQCTYSQR
FVPTPEKSVGNPEFARMRLROARNFSAARDLHKNAEMWCFRCYLSIKLALIAAD
YAVRGSQDKVPTALQKIEEYSOOLBGLNVDHVTLEAYGVDSLKTRYPDLLPPEQI
PNDRTSEVAMVMECTACIIIEKLENFMQKV"

gene

CDS

BASE COUNT 1395 a 734 c 847 g 1342 t
ORIGIN

Query Match 100.0%; Score 593; DB 9; Length 4318;
Best Local Similarity 100.0%; Pred. No. 1.3e-308;

Matches 593; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 acatctatggttacaggcttcctggttgatgaagatagcaagcaaacctcaaatggt 60
|||||
DB 3650 ACATCTTATGTTTACAGGCTTCCTGTTGATGAAAGATAGCAAGCAAACTCAAAATGGT 3709
QY 61 ggcagttctatcacagctgttagtatgtttctcgaaactgctgcgaagaacatt 120
|||||
DB 3710 GGCAGTTCATTATACAGATGTTAGTATGTTCTGGAACCTGCTGCCAAGCAACATT 3769
QY 121 tattaactgttagaactgcttctatgttctgtgtgtacataattccacaatgttata 180
|||||
DB 3770 TATTACTGTTAGAACACTGCTTATGTTTGTGTACATATTTTCCAAATGTTATA 3829
QY 181 attatatagttgtgtgaacagagatgcaactcttctgtgtctaaagtgctgcagttaa 240
|||||
DB 3830 ATTATATAGTGTGTGTAAGAGAGATCAATCTTTGTTGTCTAAAGGTCTGCACGTTAA 3889
QY 241 aaaaaaacacaccttcttccaatattgcatgttagtgagttttttaacttaaaac 300
|||||
DB 3890 AAAAAAACAACCTTTCTTTCAATATGGCATTGAGTGAGTTTATTAACCTTAATAAAC 3949
QY 301 atcaaaaattgttaaaatcatgtgttatctagtagttataaattacggcttataatc 360
|||||
DB 3950 ATCAAAAATTGTTAAATCATGTTGTATCTAGTATTATATATTCGGCTTATATTTTC 4009
QY 361 cccatgaatgatcagaactgcaattcaatcattgttctgcgcactgtcttactct 420
|||||
4

DB 4010 CCCATGAATGATCAGACGACATTTAATTCATGTTTGTCTGCCATGCTTCTTACTTT 4069
QY 421 aacatacttcttgcagaatgtaaaaggtaagataatagttatataagtgctcg 480
|||||
DB 4070 AACATATTTCTTTTGCAGAAATGTAAGGTAAGATTAATTAAGTTATTAAGGTCCTCG 4129
QY 481 ctgtaaatgcttaaatataacttctatgcaatlaagggtcttacagaacatgttgaacct 540
|||||
DB 4130 CTGTAAATGATGCTTAATATATCTTATGCAATTAAGGCTTACAGAAACATGTTGAACCTT 4189
QY 541 ttttctcttatttggaataaggaatgtttgacctccacatttatgctt 593
|||||
DB 4190 TTTTACTTTTATTTGGGAATTAAGGAATGTTTGACCTCCACATTTATTTGCTT 4242

RESULT 2

AX119931
LOCUS AX119931 12793 bp DNA linear PAT 11-MAY-2001
DEFINITION Sequence 1 from Patent WO0129266.
ACCESSION AX119931
VERSION AX119931.1 GI:14036678
KEYWORDS
SOURCE
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 12793)

AUTHORS

Hudson,T.J., Engert,J. and Richter,A.

TITLE

Identification of areas mutations and methods of use therefor

JOURNAL

Patent: WO 012926-A 1 26-APR-2001;

FEATURES

McGILL UNIVERSITY (CA) : Hopital Sainte-Justine (CA)
Location/Qualifiers

1..12793
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 4163 a 2256 c 2487 g 3887 t
ORIGIN

Query Match 100.0%; Score 593; DB 6; Length 12793;
Best Local Similarity 100.0%; Pred. No. 1.3e-308;

Matches 593; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 acatctatggttacaggcttcctggttgatgaagatagcaagcaaacctcaaatggt 60
|||||
DB 12201 ACATCTTATGTTTACAGGCTTCCTGTTGATGAAAGATAGCAAGCAAACTCAAAATGGT 12260
QY 61 ggcagttctatcacagctgttagtatgtttctcgaaactgctgcgaagaacatt 120
|||||
DB 12261 GGCAGTTCATTATACAGATGTTGATGTTCTGGAACCTGCTGCCAAGCAACATT 12320
QY 121 tattaactgttagaactgcttctatgttctgtgtgtacataattccacaatgttata 180
|||||
DB 12321 TATTACTGTTAGAACACTGCTTATGTTTGTGTACATATTTTCCAAATGTTATA 12380
QY 181 attatatagttgtgtgaacagagatgcaactcttctgtgtctaaagtgctgcagttaa 240
|||||
DB 12381 ATTATATAGTGTGTGTAAGAGAGATCAATCTTTGTTGTCTAAAGGTCTGCACGTTAA 12440
QY 241 aaaaaaacacaccttcttccaatattgcatgttagtgagttttttaacttaaaac 300
|||||
DB 12441 AAAAAAACAACCTTTCTTTCAATATGGCATTGAGTGAGTTTATTAACCTTAATAAAC 12500
QY 301 atcaaaaattgttaaaatcatgtgttatctagtagttataaattacggcttataatc 360
|||||
DB 12501 ATCAAAAATTGTTAAATCATGTTGTATCTAGTATTATATATTCGGCTTATATTTTC 12560
QY 361 cccatgaatgatcagaactgcaattcaatcattgttctgcgcactgtcttactct 420
|||||
DB 12561 CCATATGATGATGAGAACGACATTTAATTCATGTTTGTCTGCCATGCTTCTTTACTTT 12620
QY 421 aacatacttcttgcagaatgtaaaaggtaagataatagttatataagtgctacg 480
|||||

Db 12621 AACATATTTCTTTGACAGATGTAAAGTAATGATTAATGATTATTAAGTGTACTGG 12680
Oy 481 ctgtaaatgaagcacaataatacttatgcaattaaggcttacagaaactttgaactt 540
Db 12661 CTGTAAGAGATGCTAAATATCTTATGCAATTAAGGGCTTACAGAACATGTGAACTT 12740
Oy 541 ttcttactttatgggaataaggaaatgttcgacctcacattattgctt 593
Db 12741 TTTTACTTTTATTTGGCAATGATTTGCACTCCACATTTTATTTGCTT 12793

RESULT 3
AF193556 12793 bp DNA linear PRI 07-FEB-2000
LOCUS AF193556
DEFINITION Homo sapiens saccin (SACS) gene, complete cds.
ACCESSION AF193556
VERSION AF193556.1 GI:6907041
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 12793)
Engert,J.C., Berube,P., Mercier,J., Dore,C., Lepage,P., Ge,B.,
Bouchard,J.P., Mathieu,J., Melancon,S.B., Schalling,M.,
Lander,E.S., Morgan,K., Hudson,T.J. and Richter,A.
ASACS, a spastic ataxia common in northeastern Quebec, is caused
by mutations in a new gene encoding an 11.5-kb ORF
Nct. Genet. 24 (2), 120-125 (2000)
20120709
2 (bases 1 to 12793)
Engert,J.C., Berube,P., Dore,C., Lepage,P., Ge,B., Hudson,T.J. and
Richter,A.
Direct Submission
Submitted (08-OCT-1999) Genome Centre, Montreal General Hospital,
1550 Cedar Ave., Montreal, QC H3G 1A4, Canada
FEATURES
SOURCE location/Qualifiers
1. 12793
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="13"
/map="between D13S232 and D13S292"
1. 12793
/gene="SACS"
/product="saccin"
1. 12793
/gene="SACS"
77. 11566
/gene="SACS"
/note="molecular chaperone"
/codon_start=1
/product="saccin"
/protein_id="AAF31262.1"
/db_xref="GI:6907042"
translation="MNTFWGRELIVQWYPPDENRNHPSVLMKWKNNIYHSEPTL
TIDEMPLIRPTLEGGQVETLRLRPLSVLTIDDESAQLPEFLADIKYRGCEPL
KKIDASIQHPLIKYIHSPISPAVLOIMKRPLOKLNQITSLPHTKDALRKLASL
TDSSEKRIITQELAIRKINHSSDQGISSTKLKGGVLAHHTAKLPADLRLSLISVD
SSDEATIRLANMLKIEQIKTTSCILKLVKLEENAFKHEETQMLMLNIRWYISNIPAS
NPVLEWITPLKFTIOISQEQMVSAGLEPDIKLVKLEENAFKHEETQMLMLNIRWYISNIPAS
HSLEQIGLKNBASLKEKDVQVAKTEALDQVAPDDVILKRLKTLIVNKKHTLL
OSSRGKMTLKIKIWPVACKERPVPYPSILYWKDGLCNLCARPDMCDGHALIISSLP
LVESIHYNLEKALGIFTKPSLSAVLAKHFKIVDMYTSKTSDEDEYVFOHILLEYGF
MHDLNKGKDSFALKFPWWTGKFCPLQAVIRPIHDLQLYFHLNVPETAKFQ
LFPVGSIEELTSDHISMVLOKITYLSDQDSQESQONLHMLNIRWYISNIPAS
MTPVPIHHSKNPSKILMKPIHECCYCDIKDLDNLEDEVEDIILVHEDIIPKTAE
WLKVPCLSTRLINENMGFSGORREPLIVRIKILLEYVPSVDIFKELLIONADANA
TSCSLIDMRNMDIRENILDRCMAACHGALMFPNNSSPDSFVNITRGLSELRKS
EVDKVGKFGIGFNSVYHITDIPITMSKEFIMFMPNINHSKIKDSNPKIKMWSK
QQRKLRFPMQKPEFIDVFGQPLVLEAPSYGLTERFSRLIQDAAKSVESSTCY
NTADIVSLVDEFSICGHRILIFTQSVSMYLYKLIETNPSLQDQVVIKMKCSSK
ALNTPVLSVLEAKLMTKTCSSNNKLPISDEPKSSCIIQITVEEFHVRILADLQSP
LFRGPDPPALALFEMAKSGSKRPSDELQKTVBCTTILWLCCTMDTEALKFSLSSESG

RRLGLVPCGAVGVOLSEIODOKMTVRKHIGFVFCYLPRLIKTGLPVIHNGCAVTSNR
KEIMTKDPTKGMWNTPEMRHYVKAIVLOVLTSLRDLATSGELMDQTYVAWPPDPLVHD
DESVICQGFTEIDIAHGKGLKELTKVFSGGSTVSRKNNRFLDDISLTKRVDGSAKFIJF
LKYLKTKGSKMLCAVELPSSVKGLEEEAGCQIILENTFSKFPFSPFNIOETDA
ELRPDLMITFVLENEVDEFSGLRVTPCIPSLSEBHPILVPSRLIHPGGRVAKLIDRS
GRFPYGSTOYDLPNIIILIKVQLGMAADQILMDMLBRVAVSAVAKISDHYAACLRSS
IILSLIDEKTIKIRPRAKDFAKYQOTIRLPFLKPLPGPFSIDMKMGNSFKPMTNAAAD
LYTAHEODIYCLLOPIINENSHSFRGCGSVLAKVEFLGLKRTYVLIYNQLEKAVK
SVDDGITLYOENITNACYKYLHEALMONEITKMSITDKLFPSTLVBNAIVDSKYS
FHLNEAPLYLOLPNKYKNNRFLPEFVGVRSQCTVEDPALVLESIDQEGKQITVE
ENFOLCRRIISSEGIMSLIREKKOFECEKNGKILPDTNMLLPAKSIKCYDCPMWY
KDTVTKYCHADIPREVAVKLGAAYRKHAKLERVANSYCFITGTEFQKELTRKIS
ILNAVPESEKMLKELLONADAKATELCIPDPHPHVDRIEDDKMAPIDGALCVN
NQPTEEDVAGIOLGKGTGKGNRYKGOVQICGNVSYHITDCSFTSNDICITDIP
HARTAPATISIPGMRFLDADRTFQPSVLDLILGHEFLDKCTMRPRLRABEA
KAVSEISSVPASDRWONLDBLKRSDGAELELFLNMRKISICELDSTGALANLYSVK
GKITDGRKLRKROFNASVIDSVTKRRLQKLDIPVOOITYMTDDEBQNLITWILCNS
GFSSEMEKYSKVISAHKNODITLPEPGVAACTHNNKKPHRACFPILSLETJLPH
VNGHFALDSARNRMLRDNGVGRSDNNNSLMTPLIAPAYELLIOLKRRYFPGSDPT
LSVQNTPIHYVKDTLKKLSFPYRNLDLOPDICYLKAIXNCTIHDMLRILPYVA
PNIDGSHSAVITITWIMSTSNKTRPEFDVLDODEQHLNADYNTTRKTVAMNY
RLKHLLEIGFNLYNCDEANLHCLIDADIPIVSYVPADIRSFMTFSSPDNRCH
GKLEPCRLQQTNLKLEHSLKLLVDYCFDPAEENEVEGLPLITLDVLQTFDARKRK
FLTYHHELISRKDLFMMNTYLYKYSNILLNCKVAKVFDISFADLSSVLPREKTKS
CTKWKDNFASRSMKNAHFTSESVSVKEDQDEETKPFEDIVYDITKDMALLPGKFTV
SANDIVYBEGDVLPLSLMNAVPPNMOGSKVPRALMKAGCTOLALNKICGSAFAP
LISCHTANIESPISILALAHVWQTSFRRAKLVLENOFEALMTFNCNLNLMSDOI
KIILSLPCYISISGRVYSIGFGTCVYVLTNISPABEAKWQOSSSAFLEKRIHLKEI
YEVIGCVPDVDELYVLAHLPLKILENLYDAKLEHILYLNKLSAEELSEKLEDPFK
LESLLIHDANSRLKQAKHFYDRTVRYVEVWLPEKLPJNDLFKLEKRLKHNYE
MTSWVEFLRNGLKYLISOOLOFAKEISVRANTENMSKETLONTYDILHHTFOER
MDLSGNFLKELSLIIFPLCEPRABAEIRRHPOQVQENYTPILIKNGAOVNFPRKQ
DVLQMLTSCPTLEPKATPLSTKEQESDSDLPQQLQVLMMLVNLNDLPDKVINNC
RNICNITLDEEMVAKTRAKVLRISYEFLLSAEKRPFRQLRGVAVWJEDKGLKPE
VINLEYESDEPKYLYKLPLELGFPHOLFKLGEDIDISTQYEVVLSIRKNSGEO
LDPMEMTRVKKRVYGLFRSLDQNSVKRSGEDENYRDLALYPSODGRVYSILVFD
APHYSKLOGINIGOMLVDSOCYLGMDHGHTLMLIPKLPRLSLISLEQIDE
ETPVCCPGALCSQGRLOLLSSEQITGILRIKMHENMAFLANEKARLCKALR
EGLKVSCEFERLOTLVRKGFPIPHSHSEFAIRKRGNAVLTLLTQDSKDIPLFL
ALAMTLKSAENDILSDVSYLAMIGCNDIRIKIGRLDSGLGVKVSSEPSKELDPMT
IPPAEIHYYTLMDPMNAVYREGVYGVLDVADGSGDISQYPTTYAIIYVEVEDAD
NNSFLKITYQIDIDISEYKIVSLDLYKAGSPOSSSDPASPSTPSPSEFLPLGLR
SIPLPFGSKRESHTSKHOSPKLKVNSLRLPLEKTVSYVAOAKLPESEKTIIRRL
YIKWHPDKPENHDIANDVPHLONEINRELKQAFLOMDRABRRTFESASRQED
KYSFORITYSNWDATSHKSEKQOQNEKCPSPSGQTYSQRFYVPLPTFKSGNVEAR
RWLROARANEFAARNNDLHKNAENWCERYLSTLALIAADYAVARGSKDQVCKPTALA
QKIEEVSQOLEGLTNDVHTLEAVGCVDSILKRYPDLLPFPPIPNDRFTSEVAMRYMECT
ACIIRKLENEFMQKV"

BASE COUNT 4163 a 2256 c 2487 g 3887 t
ORIGIN

Query Match 100.0%; Score 593; DB 9; Length 12793;
Best Local Similarity 100.0%; Pred. No. 1.3e-308;
Matches 593; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 acaattcttgttacagcttcctcgttgatgaagaatgcaacggaaaaactcaaaatgct 60
Db 12201 ACATCTTATGTTTAAACGCTTCCTGTTGATGAGAATGATGCAACGAAATCTAAATAGT 12260
Oy 61 ggcagctttcttaccagttgttagtatgtttctggaaacagcttgcaagaacaactt 120
Db 12261 GCCAGTTCTTATTAACAGTGTGTAATGATGTTTGGGAAACGCTTGCCAAACACACTT 12320
Oy 121 tattacgttggaacacttgcttaatgtttgtgtgtacatacttccacaatgtctta 180
Db 12321 TATTACGTGTAGAACACTTGCTTATGTTGTGTGTCATATTTTCCAAATGTTATA 12380
Oy 181 attataatgtagtggatgaacagatgcaatctttgtgtcttaaggctgctgagcttaa 240
Db 12381 ATTTATATAGTGTGTTGAACGATGCAATCTTTGTTGCTTAAGGTGCTGACGTTAA 12440
Oy 241 aaaaaaacacactttcttccaatgcatgtagtgagtggttttcttaactttaaaac 300

Db	12441	AAAAAAACACCTTTCTTTCATATGCGATGATGAGAGTTTTTTAACTTTAAAAAC	12500
QY	301	atcaaaatctgtaaaatcatctgltactagttatataatcagcgttataatc	360
Db	12501	ATCAAAATTTGTTAAATCAATCTGTATCTATAGTATTAATATATCGGCTATATATTC	12560
QY	361	cccatgaatgatacgaactgacattatctatctgtctgcgcgaatgctcttcaactt	420
Db	12561	CCCATGAATGATGACAAACATGACATTTAAATCTATCTGTCTGCGCATCTCTTTACCTTT	12620
QY	421	aacatattctcttcgagaatgtaaaagtgtaaaatagttatataaagtgtaacg	480
Db	12621	AACATATTTCTTTCCAGATGTAAAAAGGTATGATTAATTAATATTAAGTGTACTGCG	12680
QY	481	ctgtaaatgtagctaaatatatacttctgcaatlaagggcttcagagaacatgltgaaact	540
Db	12681	CTGTAAATGATGCTAAATATATCTTATTCACANTTAAGGGCTTACAGAAACATGTTGAACCTT	12740
QY	541	ttttactacttcatctggagataaagaatcttgcacccctcacattttatgctt	593
Db	12741	TTTTTACCTTTTATTTGGGATTAAGGATGTCTTGGACCTCCACATTTTATTTGCTT	12793

RESULT	4
AL157766/c	
LOCUS	92693 bp
DEFINITION	DNA linear PRI 11-APR-2001
	Hunan DNA sequence from clone Rp11-40020 on chromosome 13q12.11-12.2, complete sequence.

ACCESSION	AL157766	
VERSION	AL157766.9	GI:13620292
KEYWORDS	HTG.	
SOURCE	human.	
ORGANISM	Homo sapiens	

REFERENCE
AUTHORS
TITLE
JOURNAL
Submitted (11-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
Direct Submission
Tromans, A.
1 (bases 1 to 92693)
Mammalia; Eutheria; Primates; Carnivora; Hominoidea; Homo.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
On Apr 12, 2001 this sequence version replaced q1:12709868.

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality > 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone, and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em1, EMBL; Sw1, SWISSPROT; Tr1, TREMBL; Wp1, WORMPEP; information on the WORMPEP database can be found at

http://www.sanger.ac.uk/projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 13, constructed by the Sanger Centre Chromosome 13 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/chr13> RPL140O20 is from the library RCGT-11.1 constructed by the group of Pletier de Jong. For further details see <http://www.chori.org/bacpac/home.htm> VECTOR: pBACe3.6

IMPORTANT: This sequence is not the entire insert of clone RP11-40020. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.

- The true left end of clone RP11-760M1 is at 92594 in this sequence.

```

FEATURES      The true right end of clone RP11-72P19 is at 100 in this sequence.
source        location/Qualifiers
              1..92693

```

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="13"
/map="q12.11-12.2"
/clone="RP11-40020"
/clone_1kb="RP11-11.1"
2390..2485
repeat_region
  /note="MIR repeat: matches 81..192 of consensus"
  2562..2673
repeat_region
  /note="MIR repeat: matches 68..183 of consensus"
  3895..4201
repeat_region
  /note="Alu repeat: matches 3..308 of consensus"
  5122..5397
repeat_region
  /note="MER46C repeat: matches 1..286 of consensus"
  18986..19294
repeat_region
  /note="Alu repeat: matches 1..310 of consensus"
  19644..19873
repeat_region
  /note="MER46A repeat: matches 1..235 of consensus"
  20613..20912
repeat_region
  /note="AluSg1 repeat: matches 1..300 of consensus"
  23342..23651
repeat_region
  /note="AluSg1 repeat: matches 1..309 of consensus"
  24769..24891
repeat_region
  /note="L2 repeat: matches 2554..2682 of consensus"
  25871..26011
repeat_region
  /note="L2 repeat: matches 2356..2495 of consensus"
  26033..26109
repeat_region
  /note="L2 repeat: matches 2601..2688 of consensus"
  26245..26344
repeat_region
  /note="L2 repeat: matches 2154..2255 of consensus"
  26938..27096
repeat_region
  /note="MIR repeat: matches 3..175 of consensus"
  27150..27653
repeat_region
  /note="L2 repeat: matches 1063..1644 of consensus"
  28522..28891
repeat_region
  /note="THE1B repeat: matches 1..364 of consensus"
  29447..29834
repeat_region
  /note="L1ME3A repeat: matches 5787..6164 of consensus"
  36098..36415
repeat_region
  /note="AluSx repeat: matches 1..308 of consensus"
  37202..37414
repeat_region
  /note="MIR repeat: matches 22..262 of consensus"
  37963..38254
repeat_region
  /note="AluSg repeat: matches 9..301 of consensus"
  38703..39008
repeat_region
  /note="AluSg repeat: matches 1..306 of consensus"
  39790..40093
repeat_region
  /note="AluSx repeat: matches 1..304 of consensus"
  40126..40416
repeat_region
  /note="AluSg repeat: matches 1..292 of consensus"
  40444..40733
repeat_region
  /note="AluSg repeat: matches 1..292 of consensus"
  41332..41405
misc_feature
  /note="Single clone region. Assembly confirmed by restriction digest data."
  41541..41788
repeat_region
  /note="AluSg repeat: matches 1..248 of consensus"
  44790..45101
repeat_region
  /note="AluSg repeat: matches 1..313 of consensus"
  45281..45312
repeat_region
  /note="13 copies 4 mer tgtt 888 conserved"
  45899..46206
repeat_region
  /note="Alu repeat: matches 1..307 of consensus"
  46754..47052
repeat_region
  /note="Alu repeat: matches 1..298 of consensus"
  47067..47365
repeat_region
  /note="Alu repeat: matches 1..299 of consensus"
  47477..47873
repeat_region
  /note="L1MA10 repeat: matches 5950..6322 of consensus"
  47889..48229

```


repeat_region /note="AluX repeat: matches 1. .312 of consensus"
49168. .49212
/note="Alu repeat: matches 85. .126 of consensus"
repeat_region 49620. .49693
/note="L2 repeat: matches 1685. .1757 of consensus"
repeat_region 50704. .51032
/note="AluSc repeat: matches 1. .308 of consensus"
misc_feature 52204. .53009
/note="CpG island"
evidence=not_experimental
repeat_region 53978. .54137
/note="AluX repeat: matches 1. .160 of consensus"
repeat_region 54179. .54511
/note="LMB6 repeat: matches 5822. .6172 of consensus"
repeat_region 55685. .55949
/note="AluX repeat: matches 1. .305 of consensus"
repeat_region 57331. .57390
/note="30 copies 2 mer ga 75% conserved"
repeat_region 57357. .57392
/note="9 copies 4 mer gaga 91% conserved"
repeat_region 57753. .57930
/note="MIR repeat: matches 82. .262 of consensus"
repeat_region 58260. .58389
/note="MIR repeat: matches 2. .153 of consensus"
repeat_region 58564. .58611
/note="24 copies 2 mer ca 93% conserved"
repeat_region 59350. .59533
/note="AluSq repeat: matches 129. .313 of consensus"
repeat_region 59992. .60223
/note="AluJo repeat: matches 85. .299 of consensus"
repeat_region 61036. .61144
/note="L2 repeat: matches 2581. .2696 of consensus"
repeat_region 62008. .62187
/note="TIGER1 repeat: matches 2238. .2418 of consensus"
repeat_region 62188. .62316
/note="Alu repeat: matches 1. .129 of consensus"
repeat_region 62330. .62363
/note="Alu repeat: matches 261. .294 of consensus"
repeat_region 62362. .62565
/note="TIGER1 repeat: matches 1586. .1787 of consensus"
repeat_region 62566. .62865
/note="AluSq repeat: matches 1. .302 of consensus"
repeat_region 62866. .64385
/note="TIGER1 repeat: matches 46. .1586 of consensus"
repeat_region 64386. .64694
/note="Alu repeat: matches 1. .306 of consensus"
repeat_region 64695. .64713
/note="TIGER1 repeat: matches 29. .46 of consensus"
repeat_region 65068. .65395
/note="LIPB2 repeat: matches 5405. .5733 of consensus"
repeat_region 65396. .65569
/note="Alu repeat: matches 136. .309 of consensus"
repeat_region 65571. .65640
/note="LIPB2 repeat: matches 5728. .5791 of consensus"
repeat_region 65696. .65717
/note="11 copies 2 mer ta 100% conserved"
repeat_region 65725. .66096
/note="LIPB2 repeat: matches 5789. .6155 of consensus"
repeat_region 66371. .66410
/note="10 copies 4 mer tgcg 82% conserved"
repeat_region 67586. .67886
/note="AluB repeat: matches 1. .299 of consensus"
repeat_region 69748. .69930
/note="MIR repeat: matches 6. .248 of consensus"
repeat_region 70957. .71267
/note="Alu repeat: matches 1. .311 of consensus"
repeat_region 71279. .71413
/note="MER21B repeat: matches 548. .680 of consensus"
repeat_region 71411. .71737
/note="MER31A repeat: matches 47. .485 of consensus"
repeat_region 71780. .72075
/note="AluX repeat: matches 1. .295 of consensus"
repeat_region 72145. .72256

repeat_region /note="MER31-internal repeat: matches 42. .175 of consensus"
72454. .72865
/note="MER31-internal repeat: matches 332. .739 of consensus"
repeat_region 72873. .73249
/note="MER31-internal repeat: matches 883. .1261 of consensus"
Query Match 100.0%; Score 593; DB 9; Length 92693;
Best Local Similarity 100.0%; Pred. No. 1.3e-308; Mismatches 0; Gaps 0;
Matches 593; Conservative 0; Mismatches 0; Indels 0;
OY 1 acatctatgttcaagcctcgtttgatagaagatagacaaggaacacaaatggt 60
|||||
Db 6709 ACATCTTATGTTTACAGGCTTCTGTTGATGAAGATAGCAGCGAAATCAAAATGGT 6650
|||||
OY 61 ggcagttcttattaccagttgttagtattgttctggaactgttgcgaagaacatt 120
|||||
Db 6649 GGCAGTTCTTATACAGTGTGTAGTATGTTCTGMAACTGCTGCCAAGACAACTT 6590
|||||
OY 121 tattaactgttagaacactgttctatgttgtgtgtacatatctccacaatgtcata 180
|||||
Db 6589 TATTACGTGTAGAACACTTGTCTTATGTTGTGTACATATTTTCCACAAATGTATA 6530
|||||
OY 181 attatatagtgtgttgaagaagatgcaactcttctgtgctaaaggtgtgagttaa 240
|||||
Db 6529 ATTATATAGTGTGTGAAACGAGATGCAATCTTTGTGTCTAAGGTGCTGCAGTTAA 6470
|||||
OY 241 aaaaaaacaaccccttcttcaatataatgcatgcatgtgagttcttcaattaaaac 300
|||||
Db 6469 AAAAAAACAAACCTTTCTTCAATATGCGATGTAGTGGAGTTTAACTTTAAAC 6410
|||||
OY 301 atcaaaaatgttaaaatcaatgtgtatctagtagtattataatcgcgttatctc 360
|||||
Db 6409 ATCAAAAATGTTAAATCATGTGTATGTATGTATGTTAAATATCGGCTTATATTC 6350
|||||
OY 361 cccatgaatgaatgaactgaacttcaatcaatgttgcgcgcagcttcttacttt 420
|||||
Db 6349 CCCATGAATGATCAGAACTGACATTTATTCATGTTGTCTCGCAGTCTTACTTT 6230
|||||
OY 421 aacatattcttcttgcagaatgtaaaagtaatataatattgattataaagtgtactg 480
|||||
Db 6289 AACATATTTCTTTCGAGAAATGTAAGGTATGATATATGTTATATATAGTGTACTGG 6230
|||||
OY 481 ctgtcaatgtatgctaaatataacttataatgaatgaaggtctacgaacaatgtaact 540
|||||
Db 6229 CTGTAAATGATGCTAAATATATCTTATGCAATTAAGGCTTACGAACATGTTAAACTT 6170
|||||
OY 541 ttctattttattgggaataaggaatggttgcacccacatttttgcctt 593
|||||
Db 6169 TTTTACTTTTATGGGAATPAGGAATGTTGCACTTCACATTTTATTTGCTT 6117
|||||
RESULT 5
AC079761 99819 bp DNA linear HTG 10-SEP-2000
LOCUS
DEFINITION Homo sapiens chromosome UNK clone RP11-143G17, *** SEQUENCING IN
PROGRESS ***, 44 unordered pieces.
ACCESSION AC079761
VERSION AC079761.1 GI:10047966
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 99819)
AUTHORS Waterston, R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 99819)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (10-SEP-2000) Genome Sequencing Center, Washington


```
misc_feature /note="assembly_name:Contig57"
28879..30893
/note="assembly_name:Contig58"
misc_feature /note="assembly_name:Contig58"
30994..32460
/note="assembly_name:Contig59"
misc_feature /note="assembly_name:Contig59"
32561..33984
/note="assembly_name:Contig60"
misc_feature /note="assembly_name:Contig60"
34085..35285
/note="assembly_name:Contig61"
misc_feature /note="assembly_name:Contig61"
35386..37184
/note="assembly_name:Contig62"
misc_feature /note="assembly_name:Contig62"
37285..39172
/note="assembly_name:Contig63"
misc_feature /note="assembly_name:Contig63"
39273..40874
/note="assembly_name:Contig64"
misc_feature /note="assembly_name:Contig64"
40975..42893
/note="assembly_name:Contig65"
misc_feature /note="assembly_name:Contig65"
42994..44384
/note="assembly_name:Contig66"
misc_feature /note="assembly_name:Contig66"
44485..45999
/note="assembly_name:Contig67"
misc_feature /note="assembly_name:Contig67"
46100..48669
/note="assembly_name:Contig68"
misc_feature /note="assembly_name:Contig68"
48770..50798
/note="assembly_name:Contig69"
misc_feature /note="assembly_name:Contig69"
50899..52809
/note="assembly_name:Contig70"
misc_feature /note="assembly_name:Contig70"
52910..55127
/note="assembly_name:Contig71"
misc_feature /note="assembly_name:Contig71"
55228..58087
/note="assembly_name:Contig72"
misc_feature /note="assembly_name:Contig72"
58188..61004
/note="assembly_name:Contig73"
misc_feature /note="assembly_name:Contig73"
61105..64185
/note="assembly_name:Contig74"
misc_feature /note="assembly_name:Contig74"
64286..67105
```

Query Match	100.0%	Score 593;	DB 2;	Length 99819;
Best Local Similarity	100.0%	Pred. No. 1,36-308;		
Matches 593;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY 1	acattctatggttaagcctctctcttcttgataagaatagcaacgnaaacctcaaatgtg	60		
Db 61122	ACATTTTATGTTTACAGCGCTTCCTGTGATGAAGATACCAACGGAAACTCAAAATGGT	61181		
QY 61	ggcagttcttattacacagtggttgaatgtgtcttcggaacgctgtccgaacaaact	120		
Db 61182	GGCATTTCTTATTACACAGTGTGATGTTGTTTCGGAACGCTTGCCACAGCAACATTT	61241		
QY 121	tattacgtgttaagaacacttgcttctatgtgttgltgltacataattcccaaatgtata	180		
Db 61242	TATTAACGTGTGAGAACACTTGCTTATGTTGTGTGTGATACATTTTCCACAATGTTATA	61301		
QY 161	attctatagtggtgttgaaacagatgcgaactcttctgtgtctaaagtgctgcagttaa	240		
Db 61302	ATTTATATAGTGTGCTTGAAACAGATGCATTTTGTGTGTCTMAAAGGTGCTGCAGTTAA	61361		
QY 241	aaaaaaaaaacactttctcttcaataatgacatgtagtgagtggttttttctaactttaaac	300		
Db 61362	AAAAAAAAAACACCTTTCTTCTTAATATGGCAATGATGGAGTTTTTTTAACTTTAAAAAC	61421		
QY 301	atcaaaacttgttaaatcatcatgtgttactagtagtattcaatcaatgcgtataatcc	360		
Db 61422	ATCAAAATTTGTTAAATACATGTGTGATATCTAGTACTTATTAATTTATCGGCTTAATTTTC	61481		
QY 361	cccatgaatgacagaaactgaacatcttaacttcatgtgttgcctgcacatgctcttcaactt	420		
Db 61482	CCCATGAATGATCAACAACAGTACATTTAATTAATCATGTTTGTCGCCATGCTTCTTACTTT	61541		
QY 421	aacataattctcttgcagaatgtaaaaggtaatgataattagttatataaagtgtacg	480		
Db 61542	AACATAATTTCTTTTCAGATATGTAATAAAGTAATGATAAATAGTTTATATAAGTGTACTCG	61601		

QY 481 cgttaaatgagtcgaacataatactcttatgcacatbaagggcacaagaacatggtgaact 540

61602 CTGTAAATGATGCTCAATAATATCTTATGCACATTTAAGGCGTTACGACACATGTTGAACCT 616

QY 541 ttttaactcttattcgggaataaggaaatglttgcacatccacatlttatgctc 593

Db 61662 TTTTATCTTATTTATGGGAAATTAAGGAATGTTGCAACCTCCACATTTTATTCCTT 61714

RESULT	6					
G36555/c						
LOCUS	G36555	418 bp	mRNA	linear	STS 31-DEC-1997	
DEFINITION	SHCC-53325	Human Homo sapiens	STS cDNA,	sequence tagged site.		
ACCESSION	G36555					
VERSION	G36555.1	GI:2734222				
KEYWORDS	STS,					
SOURCE	human.					
ORGANISM	Homo sapiens					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.					
TITLE	1 (bases 1 to 418)					
JOURNAL	Myers, R.M.					
COMMENT	Human STS (1997)					
	unpublished					

Contact: Richard M. Myers
Stanford Human Genome Center (SHGC)
Stanford University School of Medicine
Department of Genetics, M-344, Stanford, CA 94305, USA
Tel: 4157259687
Fax: 4157259683
Email: myers@shgc.stanford.edu
Primer A: GTGGAGGTGCAACATCTCT
Primer B: ACATTATTCATCATTTGTCCCG
STS size: 202
PCR profile:

Initial incubation:	95 degrees C for 10 minutes
Denaturation:	94 degrees C for 30 seconds
Annealing:	60 degrees C for 30 seconds
Polymerization:	72 degrees C for 23 seconds
PCR Cycles:	30
Thermal Cycler:	Perkin Elmer 9600
Protocol:	
Template:	25 ng
Primer:	each 1 uM
dNTPs:	each 200 uM
Amplifad Gold Polymerase:	0.07 units/ul
Total Vol:	5 ul

Buffer:

MgCl ₂ :	2.5 mM
KCl:	50 mM
Tris-HCl:	10 mM
pH:	8.3

Prepared with primer pairs derived from N46342 -- Unigene.
Location/Qualifiers

source	1. .418
/organism="Homo sapiens"	
/db_xref="taxon:9606"	
/map="13"	
/clone_1lb="Human"	
STS	89. .290
primer bind	89. .108
primer_bind	complement(268. .290)
BASE COUNT	167 a 65 c 55 g 130 t
ORIGIN	1 others

Query Match	26.1%;	Score 155;	DB 11;	Length 418;
Best Local Similarity	99.28;	Pred. No. 2.7e-72;		
Matches 255;	Conservative 0;	Mismatches 2;	Indels 0;	Gaps 0;

QY	253	ctttctcttccaatagcgatgtagtggaggttttttaactttaaaaacatccaaaattgt	312
Db	418	CTTTCTTTCAATANGGCGATGAGTGGAGTTTTTTAACTTTAAAAACATCAAAANTTGT	359
QY	313	taaaatcatctgltatcatagtagitttaaatlalcggtcattattcccatgtaatgat	372
Db	358	TAAATCATGTGTGTATCTAGTACTTTTAAATTANCGGTATATTTCGCCATGATGAT	299
QY	373	cgaacttcaactttaattcatggttgcgcgaatcttctaactttaacaatttctt	432
Db	298	CAGACTCAACTTTTATATCATGTTGTGCCGCAATGCTCTTACTTTAACTATTTCCT	239
QY	433	ttgcagaaatgtcaaaagtgatataatagttatataagtgtaacgctgtcaaatgaty	492
Db	238	TTGCAGAAATGTAAAAAGTAAATCATATATTAGTTATATATAAGTACTGGCTGTAATATGANG	179
QY	493	ctaaataactttaatgc	509
Db	178	CTAAATATACTTTTATGC	162

RESULT	7
AB056815	
LOCUS	
DEFINITION	AB056815 3389 bp mRNA linear PRI 14-MAR-2001
ACCESSION	Macaca fascicularis brain cDNA clone:Qf1A-15307, full insert sequence.
VERSION	AB056815
KEYWORDS	AB056815.1 GI:13365931
SOURCE	its (full insert sequence); oligo capping.
ORGANISM	Macaca fascicularis adult male frontal lobe left cDNA to mRNA, clone_11b:macaque brain cDNA library Qf1A clone:Qf1A-15307.
	Macaca fascicularis

REFERENCE
AUTHORS
TITLE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
Cercopithecinae; Macaca.
1 (sites)
Osada, N., Hida, M., Kusuda, J., Tanuma, R., Iseki, K., Hirai, M.,
Terao, K., Suzuki, Y., Sugano, S. and Hashimoto, K.
Isolation of full-length cDNA clones from macaque brain cDNA
libraries

Unpublished
2 (bases 1 to 3289)
Hashimoto, K., Osada, N., Hida, M., Kusuda, J. and Sugano, S.
Direct Submission
Submitted (09-MAR-2001) katsuyuki Hashimoto, National Institute of

COMMENT
Lab Host: TOP10
Infectious Diseases, Division of Genetic Resources: 23-1, Toyama
1-chome, Shinku-Ku, Tokyo 162-6640, Japan
(E-mail: hashi@nih.go.jp URL: <http://www.nih.go.jp/yoken/genebank/>,
Tel: 81-3-5285-1111 (ex. 2120), Fax: 81-3-5285-1181)

Vector: pME18s-FL3 (Acc. No. AB009864)
R. Site1: DraIII (CAGCTGTC)
R. Site2: DraIII (CAGCAGTC)
Description: 1st strand cDNA was primed with an oligo(dT) primer [ATGCGCCCTTTTCTTTTCTTTT]; double-stranded cDNA was synthesized using specific 5' and 3' primers and amplified by PCR. The PCR product was digested with SfiI and size selection was performed to exclude fragments <1.5kb. The SfiI-digested PCR product was cloned into distinct DraIII sites of pME18s-FL3. XhoI sites just outside the DraIII sites can be used to isolate the cDNA insert. Libraries were constructed by Sugano et al. (University of Tokyo, Institute of Medical Science). Custom primer used for sequencing
(5' end primer [CTTCTCTCTTAAGCTCGG] ;
3' end primer [GCAGCTGACCTCGACACA]).

```

FEATURES      location/qualifiers
source        1. .3289

```

```

/organism="Macaca fascicularis"
/db_xref="taxon:9541"
/clone="Qf1A-15307"
/sex="male"
/tissue_type="frontal lobe left"
/clone_lib="macaque brain cDNA library Qf1A"

```

```

/dev_stage="adult"
/not="host; TOP10; Vector; pME185-FLJ (Acc. No. AB009664);
R_Site1: DraII (CACTGTGTG); R_Site2: DraII
(CACCATGTG)"
35. .1951

```

```

/codon_start=1
/product="hypothetical protein"
/protein_id="BAB39340.1"
/db_xref="GI:13365932"
/translation="MTVDDLSQCYLGNDRGHTKIMLPQKILAPRLSLRLDDEE
TPKVCDFGALCSLOGRLQDLSSSEQFTTGILIRIKMKNDAFLANEKALIRLCKALRE
GLKVSCEKQLQTLIRVKGFPNDIPSRSETEPAFLKRENNAILYLTQHSBNDKINFLIA
LAMIATSATNLDLSITSYLIAMLCNDNIYIGELSLGKVVSSSESKLELPMPGP
IPAEIHTYLLMDPMNPNVYPSGEYGLVADACGGDLYSGYQPTYYAIIIVQEVERDADN
SSEFLGKIQYDIDIGSEYKIVSLDLYKFSPPDESSQSRDASPTSPTEFLAGLGS
IPLPFGRESHKTSKSHQSPKRLKVNLPETLKEVTSVQAMKLPSEKRTILRLY
LKMPDPKPNENHDIAVFNKHLONEITRLKQFLDNADRASRPPTSSASRQDSX
YRSRQFTYSNNGEATSHKSERDOONKKKCPPSAQQTSQRFVPTPTFKSGNYPEAR
WLRQARANFSAARNDLTKNANENWCVCYISTKLALVAADYAVAGSKDKVCPALAQ
KIEEYSOULEGLTNDVHTLPAAYGVDSLKTRYPDLLPQPIPNDRFTSEVAMRWECTA
CIILKENFVIOQKV"

```

BASE COUNT	1062 a	552 c	653 g	1022 t
ORIGIN				
Query Match	16.2%	Score 96	DB 9	Length 3289
Best Local Similarity	99.3%	Pred. No.	1.8e-40	
Matches 146	Conservative 0	Mismatches 1	Indels 0	Gaps 0

Oy	92	tttcggaaacgctcttggccaagacaacatttatttaactgttagaacacttgcttcatg	151
Db	2664	TTCTGGAACACGCTTGGCAACACACATTTTAACTGTGGAACACTGCTTATATGTT	2743
Oy	152	gtgtgtacatacttcaccaaaegtataacttatalatggtggttgacagagatgc	211
Db	2744	GGTGTATATATTTTCCACAATGTTAATTTATATGTGTGTAACAGGATGCAT	2803
Oy	212	ctttgtgtgtctaaagtgctgcagt	238
Db	2804	CTTTGTGTCTAAAGGTGCTGCAGT	2830

RESULT	8	51 bp	DNA	linear	PAT 22-JUN-2001
AX160575					
LOCUS	AX160575				
DEFINITION	Sequence 3903 from Patent WO0140521.				
ACCESSION	AX160575				
VERSION	AX160575.1				
KEYWORDS	GI:14541906				
SOURCE	human.				
ORGANISM	Homo sapiens				

REFERENCE	JOURNAL	FEATURES	SOURCE
Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi: Mammalia: Eutheria: Primates: Catarrhini: Homnidae: Homo. 1 (bases 1 to 51)			
Shimkets, R.A. and Leach, M. Nucleic acids containing single nucleotide polymorphisms and methods of use thereof	Patent: WO 0140521-A 3903 07-JUN-2001;		
Curagen Corporation (US)		Location/Qualifiers	
		1..51	

```
misc_feature      /organism="Homo sapiens"
                  /db_xref="taxon:9606"
                  26
note="1 of 2 allelic variants (3904 is other entry)
Accession number CG43924289"
BASE COUNT      15 a      8 c      9 g      20 t
ORIGIN
```

```
Query Match      8.6%: Score 51: DB 6: Length 51:
Best Local Similarity 100.0%: Pred. No. 3.5e+16:
Matches 51: Conservative 0: Mismatches 0: Indels 0: Gaps 0:
```

Oy 109 caagacacattatcattagacacttcttattgtgtgtac 159
 |||||||
 Db 1 CAAGACACATTATTAAGTGTAGAACACTTGTATGTGTGTAC 51

RESULT 9

AX160577 51 bp DNA linear PAT 22-JUN-2001
 LOCUS Sequence 3905 from Patent WO0140521.
 DEFINITION AX160577
 ACCESSION AX160577 GI:14541908
 VERSION AX160577.1
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 51)
 Shimkets, R.A. and Leach, M.
 Nucleic acids containing single nucleotide polymorphisms and
 methods of use thereof
 Patent: WO 0140521-A 3905 07-JUN-2001;
 Curagen Corporation (US)

JOURNAL

Location/Qualifiers
 1. 51

FEATURES

source
 /organism="Homo sapiens"
 /db_xref="taxon:9606"

misc_feature

26 /note="1 of 2 allelic variants (3906 is other entry)
 Accession number cg43924289"

BASE COUNT 20 a 8 c 8 g 15 t
 ORIGIN

Query Match 8.6%; Score 51; DB 6; Length 51;
 Best Local Similarity 100.0%; Pred. No. 3.5e-16;
 Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 223

taaaagtcgtgcagttcaaaaacaccttcttcaatagtcagt 273
 |||||||
 Db 1 TAAAGTGTGTCAGTAAAAAACACACTTCTTCATATGCGCATG 51

RESULT 10

AX160578 50 bp DNA linear PAT 22-JUN-2001
 LOCUS Sequence 3906 from Patent WO0140521.
 DEFINITION AX160578
 ACCESSION AX160578
 VERSION AX160578.1 GI:14541909
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 50)
 Shimkets, R.A. and Leach, M.
 Nucleic acids containing single nucleotide polymorphisms and
 methods of use thereof
 Patent: WO 0140521-A 3906 07-JUN-2001;
 Curagen Corporation (US)

Location/Qualifiers
 1. 50

JOURNAL

Location/Qualifiers
 1. 50

FEATURES

source
 /organism="Homo sapiens"
 /db_xref="taxon:9606"

misc_feature

25. 26 /note="Nucleotide deleted between bases 25 and 26
 Accession number cg43924289"

misc_feature

26 /note="2 of 2 allelic variants (3905 is other entry)"
 Accession number cg43924289"

BASE COUNT

19 a 8 c 8 g 15 t
 ORIGIN

Query Match 5.7%; Score 34; DB 6; Length 50;
 Best Local Similarity 100.0%; Pred. No. 5.2e-07;
 Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 240

aaaaaaaaaacaccttcttcaatagtcagt 273
 |||||||
 Db 17 AAAAAAAAAACACTTCTTCATATGCGCATG 50

RESULT 11

AX160576 51 bp DNA linear PAT 22-JUN-2001
 LOCUS Sequence 3904 from Patent WO0140521.
 DEFINITION AX160576
 ACCESSION AX160576
 VERSION AX160576.1 GI:14541907
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 51)
 Shimkets, R.A. and Leach, M.
 Nucleic acids containing single nucleotide polymorphisms and
 methods of use thereof
 Patent: WO 0140521-A 3904 07-JUN-2001;
 Curagen Corporation (US)

JOURNAL

Location/Qualifiers
 1. 51

FEATURES

source
 /organism="Homo sapiens"
 /db_xref="taxon:9606"

misc_feature

26 /note="2 of 2 allelic variants (3903 is other entry)
 Accession number cg43924289"

BASE COUNT 14 a 8 c 9 g 20 t
 ORIGIN

Query Match 4.2%; Score 25; DB 6; Length 51;
 Best Local Similarity 100.0%; Pred. No. 0.037;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 109

caagacacattatcattagtcagt 133
 |||||||
 Db 1 CAAGACACATTATTAAGTGTAG 25

RESULT 12

AX160574 50 bp DNA linear PAT 22-JUN-2001
 LOCUS Sequence 3902 from Patent WO0140521.
 DEFINITION AX160574
 ACCESSION AX160574
 VERSION AX160574.1 GI:14541905
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 50)
 Shimkets, R.A. and Leach, M.
 Nucleic acids containing single nucleotide polymorphisms and
 methods of use thereof
 Patent: WO 0140521-A 3902 07-JUN-2001;
 Curagen Corporation (US)

Location/Qualifiers
 1. 50

JOURNAL

Location/Qualifiers
 1. 50

FEATURES

source
 /organism="Homo sapiens"
 /db_xref="taxon:9606"

misc_feature

25. 26 /note="Nucleotide deleted between bases 25 and 26
 Accession number cg43924289"

misc_feature

26 /note="2 of 2 allelic variants (3901 is other entry)"
 Accession number cg43924289"

BASE COUNT

12 a 8 c 3 g 27 t

ORIGIN

Query Match 4.0%; Score 24; DB 6; Length 50;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 acatctatgttaccagcttct 24
|||||
Db 27 ACATCTATGTTTACAGGCTTCT 50

RESULT 13

AX160573 51 bp DNA linear PAT 22-JUN-2001
LOCUS Sequence 3901 from Patent WO0140521.
DEFINITION AX160573
ACCESSION AX160573
VERSION AX160573.1 GI:14541904
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE 1 (bases 1 to 51)
Shinketsu, R.A. and Leach, M.
Nucleic acids containing single nucleotide polymorphisms and
methods of use thereof
Patent: WO 0140521-A 3901 07-JUN-2001;
JOURNAL Curagen Corporation (US)

FEATURES

source Location/Qualifiers
1..51
/organism="Homo sapiens"
/db_xref="taxon:9606"

misc_feature /note="1 of 2 allelic variants (3902 is other entry)
26
Accession number cg43924289"

BASE COUNT 12 a 8 c 3 g 28 t
ORIGIN

Query Match 4.0%; Score 24; DB 6; Length 51;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 acatctatgttaccagcttct 24
|||||
Db 28 ACATCTATGTTTACAGGCTTCT 51

RESULT 14

AL670276 149128 bp DNA linear HTG 30-JAN-2002
LOCUS Mus musculus chromosome 4 clone RP23-115D21, *** SEQUENCING IN
DEFINITION
PROGRESS***, in unordered pieces.

ACCESSION AL670276
VERSION AL670276.2 GI:18477095
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE house mouse.
ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
TITLE 1 (sites)
AUTHORS Burton, J.
JOURNAL

Submitted Direct Submission
Cambridge, MA 02138, USA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
On Feb 1, 2002 this sequence version replaced gi:18307375.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk

Project Information

Center project name: bm115D21
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 146474 bases at least Q40
Consensus quality: 146902 bases at least Q30
Consensus quality: 147368 bases at least Q20
Insert size: 148128; sum-of-contigs
Insert size: 154153; 3.3% error; agarose-fp
Quality coverage: 10.26x in Q20 bases; sum-of-contigs Quality
coverage: 9.86x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

FEATURES

source

Location/Qualifiers
1..149128
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="4"

misc_feature /clone="RP23-115D21"
/clone_lib="RPCT-23"
1..4701

misc_feature /note="assembly-fragment:02491
fragment_chain:1
clone_end:sp6
vector_side:left"
4802..9220

misc_feature /note="assembly-fragment:00966
fragment_chain:1"
9321..12084

misc_feature /note="assembly-fragment:00594
fragment_chain:1"
12185..17776

misc_feature /note="assembly-fragment:01529
fragment_chain:1"
17877..22373

misc_feature /note="assembly-fragment:01073
fragment_chain:1"
22474..30439

misc_feature /note="assembly-fragment:00975
fragment_chain:1"
30540..37808

misc_feature /note="assembly-fragment:00276
fragment_chain:1"
37909..117226

misc_feature /note="assembly-fragment:01606
fragment_chain:1"
117327..133040

misc_feature /note="assembly-fragment:01524
fragment_chain:1"
133141..139559

misc_feature /note="assembly-fragment:00876
fragment_chain:1"
139660..149128

misc_feature /note="assembly-fragment:01484
clone_end:t7
vector_side:right"
BASE COUNT 39357 a 33291 c 33024 g 42447 t 1009 others
ORIGIN

Query Match 3.9%; Score 23; DB 2; Length 149128;
Best Local Similarity 100.0%; Pred. No. 0.42;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 239 aaaaaaaaaaacacttcttt 261
|||||
Db 147645 AAAAAAAAAACACCTTCTTT 147667

```

RESULT 15
AC092096 336028 bp DNA linear HTG 05-FEB-2002
LOCUS Mus musculus chromosome 10 clone rp23-39k4, WORKING DRAFT SEQUENCE,
DEFINITION 19 unordered pieces.
ACCESSION AC092096
VERSION AC092096.17 GI:18497142
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognath; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 336028)
AUTHORS Do,T. and Roe,B.A.
TITLE Mus musculus Chromosome 10 BAC Clone rp23-39k4
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 336028)
AUTHORS Do,T. and Roe,B.A.
TITLE Submitted
JOURNAL Direct Submission
Submitted (20-JUN-2001) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
On Feb 5, 2002 this sequence version replaced gi:18201826.
-----
Center: Department Of Chemistry And Biochemistry
The University Of Oklahoma
Center code:UOKNOR
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 19 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
3074: contig of 3074 bp in length
3075 3174: gap of unknown length
3175 6014: contig of 2840 bp in length
6015 6114: gap of unknown length
6115 9345: contig of 3231 bp in length
9346 9445: gap of unknown length
9446 12107: contig of 2662 bp in length
12108 12207: gap of unknown length
12207 18756: contig of 6549 bp in length
18757 18856: gap of unknown length
18857 23733: contig of 4877 bp in length
23734 23833: gap of unknown length
23833 30446: contig of 6613 bp in length
30447 30546: gap of unknown length
30547 37867: contig of 7321 bp in length
37868 37967: gap of unknown length
37968 45708: contig of 7741 bp in length
45709 45809: gap of unknown length
45809 53387: contig of 7579 bp in length
53388 53487: gap of unknown length
53488 63026: contig of 9539 bp in length
63027 63126: gap of unknown length
63127 74700: contig of 11574 bp in length
74701 74800: gap of unknown length
74801 85536: contig of 10736 bp in length
85537 85636: gap of unknown length
85637 108073: contig of 22437 bp in length
108074 108174: gap of unknown length
108175 137982: contig of 29709 bp in length
137983 137983: gap of unknown length
137984 183433: contig of 45451 bp in length
183434 183533: gap of unknown length
183534 222828: contig of 39295 bp in length
222829 222928: gap of unknown length
222929 262269: contig of 39341 bp in length
262270 262369: gap of unknown length
262370 336028: contig of 73659 bp in length.

```

```

FEATURES
    source
        location/Qualifiers
            1..336028
                /organism="Mus musculus"
                /db_xref="taxon:10090"
                /chromosome="10"
                /clone="rp23-39k4"
                /clone_1lb="rp23"
BASE COUNT 101266 a 67078 c 66482 g 99351 t 1851 others
ORIGIN
Query Match 3.9% Score 23; DB 2; Length 336028;
Best Local Similarity 100.0%; Pred. No. 0.42;
Matches 23; Conservative 0; Mismatches 0; Gaps 0;
Qy 239 aaaaaaaaaaaccttcttctt 261
| | | | | | | | | | | | | | | |
Db 62506 AAAAAAAAAAACACCTTTCTT 62528
Search completed: May 22, 2002, 08:50:10
Job time: 11900 sec

```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 22, 2002, 08:36:50 ; Search time 373 Seconds

(without alignments)
2729.572 Million cell updates/sec

Title: US-09-693-205-7_COPY_12200_12792

Perfect score: 593

Sequence: 1 acatctatgtttacaggtc.....acctccacattatgtctt 593

Scoring table:

OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 1736436 seqs, 858457221 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : N_Genesed_032802:*

1: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	593	100.0	1317	22	AAS29058
2	593	100.0	12792	22	AAH20176
3	593	100.0	12793	22	AAH20174
4	593	100.0	12793	22	AAH20178
5	593	100.0	12793	22	AAH20179
6	593	100.0	12793	22	AAH20182
7	430	72.5	1387	22	AAS29132
8	51	8.6	51	22	AAI76962
9	51	8.6	51	22	AAI76964

10	34	5.7	50	22	AAI76965	Human silent SNP c
11	25	4.2	51	22	AAI76963	Human silent SNP c
12	24	4.0	50	22	AAI76961	Human silent SNP c
13	24	4.0	51	22	AAI76960	Human silent SNP c
14	21	3.5	4604	22	AAS45491	Chemically pretrea
15	20	3.4	268	20	AAV90383	EST clone DM420.
16	20	3.4	1947	22	AAH16703	Human CDNA sequenc
17	19	3.2	384	21	AAH30804	Human colon cancer
18	19	3.2	504	23	ABL26223	Drosophila melanog
19	19	3.2	2504	23	ABL26222	Drosophila melanog
20	19	3.2	2617	23	ABL21778	Drosophila melanog
21	19	3.2	5238	24	ABL32758	Human immune syste
22	19	3.2	6052	24	ABL32417	Human immune syste
23	19	3.2	6081	24	ABL33428	Human immune syste
24	19	3.2	8770	22	AAS46571	Tumour suppressor
25	19	3.2	8770	24	AAS61353	Human gene regulat
26	18	3.0	402	21	AAC59032	Human breast cell
27	18	3.0	462	22	ABA42960	Human foetal liver
28	18	3.0	462	22	ABA53380	Probe #1624 for ge
29	18	3.0	462	22	ABA23158	Human brain expres
30	18	3.0	462	22	AAK01647	Human bone marrow
31	18	3.0	462	22	AAK27093	Probe #1620 for ge
32	18	3.0	462	22	AAI11687	Probe #1675 used t
33	18	3.0	462	22	AAI32989	Probe #1604 used t
34	18	3.0	462	22	AAI01613	Human CDNA clone (
35	18	3.0	570	22	AAH10393	CDNA encoding a hu
36	18	3.0	784	21	AAAS8994	Genomic sequence #
37	18	3.0	784	21	AAAS9161	Human secreted pro
38	18	3.0	1026	22	AAS28472	Human secreted pro
39	18	3.0	1026	22	AAS28474	Human secreted pro
40	18	3.0	1037	21	AAC599721	Human secreted pro
41	18	3.0	1075	22	ABAI7718	Human nervous syst
42	18	3.0	1711	21	AAC77140	Human ORFX ORF2695
43	18	3.0	1770	22	AAH17519	Human CDNA sequenc
44	18	3.0	1972	22	AAD07834	Human secreted pro
45	18	3.0	2573	18	AAI77804	CDNA encoding vari

ALIGNMENTS

RESULT 1	
ID	AAS29058 standard; CDNA: 1317 BP.
XX	XX
AC	AAS29058;
XX	XX
DT	21-NOV-2001 (first entry)
XX	XX
DE	CDNA encoding for human DNA-binding protein #29.
XX	XX
KW	Human: DNA-binding protein; histone; chromo domain protein;
KW	chromatin organisation modifier; y-box binding protein;
KW	DNA organisation; gene transcription; malignant disease;
KW	autoimmune disorder; rheumatic disease; genetic abnormality;
KW	infectious disease; neurological disorder; gene therapy;
KW	immunomodulatory; anti-HIV; anti rheumatic; anti microbial;
KW	cytostatic; ss.
XX	XX
OS	Homo sapiens.
XX	XX
PN	WO200155162-A1.
XX	XX
PD	02-AUG-2001.
XX	XX
PF	17-JAN-2001; 2001MO-US01305.
XX	XX
PR	31-JAN-2000; 2000US-0179065.
XX	XX
PR	04-FEB-2000; 2000US-0180628.
XX	XX
PR	24-FEB-2000; 2000US-0184664.
XX	XX
PR	02-MAR-2000; 2000US-0186350.
XX	XX
PR	16-MAR-2000; 2000US-0189874.
XX	XX
PR	17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216860.
PR 11-JUL-2000; 2000US-0217487.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 14-AUG-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0237182.
PR 23-AUG-2000; 2000US-0237009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234224.
PR 25-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.

PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0250130.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
PI
XX
XX WPI; 2001-465557/50.
DR
XX P-PSDB; AAU18182.
XX
XX Nucleic acid molecules encoding human secreted chromosomal binding
PT proteins, used in preventing, treating or ameliorating a disorder, e.g.
PT Alzheimer's and Parkinson's diseases and cancers -
XX
XX
PS Claim 4; SEQ ID No 39; 561pp; English.
XX
CC The present invention relates to the isolation of novel DNA-binding
CC proteins (AAU18154-AAU18281), and cDNA and genomic sequences encoding

CC for these proteins. DNA-binding proteins such as histones, chromo
 CC (chromatin organisation modifier) domain proteins, and Y-box binding
 CC proteins may contribute to diseases resulting from aberrant DNA
 CC organisation and/or gene transcription. The sequences of the invention
 CC are useful in screening assays to identify antagonists and/or agonists
 CC that may enhance or block activities mediated by DNA-binding proteins.
 CC Blockers of DNA-binding proteins may be useful in treating disorders
 CC such as malignant diseases (e.g. cancer), autoimmune disorders
 CC (e.g. diabetes mellitus), rheumatic diseases (e.g. rheumatoid
 CC arthritis), genetic abnormalities (e.g. cystic fibrosis), infectious
 CC diseases (e.g. HIV) and neurological disorders (e.g. Alzheimer's
 CC disease). The polynucleotide sequences of the invention may also be
 CC used in gene therapy. AAS29030-AAS29157 represent cDNA sequences
 CC encoding for novel DNA-binding proteins.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at http://wipo.int/pub/published_pcl_sequences.

SQ Sequence 1317 BP; 418 A; 167 C; 243 G; 489 T; 0 other;

Query Match 100.0%; Score 593; DB 22; Length 1317;
 Best Local Similarity 100.0%; Pred. No. 1.9e-272;
 Matches 593; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 acatctatgttacaagcctctgttgaatagaacaggaacgaactcaaatggt 60
 Db 629 acatctatgttacaagcctctgttgaatagaacaggaacgaactcaaatggt 688
 QY 61 ggcagttctatcacagctgttgaatgttcttgcgaactgcttgcgaacacatt 120
 Db 689 ggcagttctatcacagctgttgaatgttcttgcgaactgcttgcgaacacatt 748
 QY 121 tattaactgttagaacacattgttattgttgcgaactgcttgcgaacacatt 180
 Db 749 tattaactgttagaacacattgttattgttgcgaactgcttgcgaacacatt 808
 QY 181 attatatagttgcgttgaacaggaatgcaatttgcgttgcgaacaggaatgca 240
 Db 809 attatatagttgcgttgaacaggaatgcaatttgcgttgcgaacaggaatgca 868
 QY 241 aaaaaaaacacaccttcttccaataatgcatgtagtgaggttttctaacttaaac 300
 Db 869 aaaaaaaacacaccttcttccaataatgcatgtagtgaggttttctaacttaaac 928
 QY 301 atcaaaaattgttaaaatcatctgtatcctagtagtitaatattatcggtataattc 360
 Db 929 atcaaaaattgttaaaatcatctgtatcctagtagtitaatattatcggtataattc 988
 QY 361 ccgatgatgatcagacatgacatttaattcattgtgtctgcgacgtcttcttactt 420
 Db 989 ccgatgatgatcagacatgacatttaattcattgtgtctgcgacgtcttcttactt 1048
 QY 421 aacataattcttcgcgaatgtcaaaaggaatgataatagttatataagttactgg 480
 Db 1049 aacataattcttcgcgaatgtcaaaaggaatgataatagttatataagttactgg 1108
 QY 481 ctgtaaatgtctaaatataacttattgcaatgaatgaaggtctacagaaatgttgaact 540
 Db 1109 ctgtaaatgtctaaatataacttattgcaatgaatgaaggtctacagaaatgttgaact 1168
 QY 541 ttttcttatttgggaataaagaatgttgcaccccaactttattgctt 593
 Db 1169 ttttcttatttgggaataaagaatgttgcaccccaactttattgctt 1221

RESULT 2

AAH20176
 ID AAH20176 standard; DNA: 12792 BP.

XX AAH20176;

DT 09-AUG-2001 (first entry)

XX Human mutated spastin nucleotide sequence SEQ ID NO:7.
 DE
 XX
 KW Human; mouse; spastin; ARSACS; chromosome 13q11; Identification;
 KW autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;
 KW neurodegenerative disease; reduced sensory nerve conduction; diagnosis;
 KW reduced motor nerve velocity; hypermyelination of retinal nerve fibre;
 KW atrophy of upper cerebellar vermis; absence of Purkinje cell;
 KW abnormal neuronal lipid storage; genetic disorder; characterisation; ds.
 XX
 OS Homo sapiens.
 OS Synthetic.
 FH Key location/Qualifiers
 FT CDS 77..6604
 FT /tag= a
 FT /product= "mutated spastin"

W0200129266-A2.

XX 26-APR-2001.

XX 20-OCT-2000; 2000MO-US29130.

XX 20-OCT-1999; 99US-0160588.

XX (UYMC-) UNIV MCGILL.
 PA (HOP1-) HOPITAL SAINTE-JUSTINE.

PI Hudson TJ, Engert J, Richter A;

DR WPI, 2001-308494/32.

XX P-PSDB; AAB97821.

XX New isolated polynucleotide, encoding spastin gene, and polypeptides,
 PT useful for diagnosing autosomal recessive spastic ataxia of
 PT Charlevoix-Saguenay disease by detecting two point mutations in spastin
 PT gene sequence -

XX Claim 1; Page -: 76pp; English.

XX The present invention describes human and mouse spastin, and mutated
 CC human spastin (autosomal recessive spastic ataxia of Charlevoix-Saguenay
 CC (ARSACS)) gene sequences (I). The spastin gene has been mapped to
 CC chromosome 13q11. (I) have neuroprotective activities and can be used in
 CC gene therapy and as a spastin polypeptide agonists. (I), their fragments
 CC or their complements can be useful for assaying the presence of a nucleic
 CC acid molecule in a sample. (I) is useful for diagnosing or aiding in the
 CC diagnosis of an early onset neurodegenerative disease in an individual.
 CC The neurodegenerative disease comprises reduced sensory nerve conduction,
 CC reduced motor nerve velocity, hypermyelination of retinal nerve fibres,
 CC atrophy of upper cerebellar vermis, absence of Purkinje cells and
 CC abnormal neuronal lipid storage. (I) can also be used to produce
 CC antisense nucleic acids, is useful as molecular weight or chromosome
 CC markers, to identify genetic disorders, as hybridisation probes or
 CC primers, as an antigen, identify and express recombinant protein for
 CC analysis, characterisation or therapeutic use, or as markers for tissues
 CC in which the corresponding protein is expressed. Diagnostic methods from
 CC the present invention can be used to identify subjects having or at risk
 CC of developing a disease or disorder associated with aberrant expression
 CC or activity of (I). The assays can be utilised to identify a subject
 CC having or at risk of developing a disorder associated with Spastin
 CC protein or spastin gene expression or activity. The present sequence
 CC encodes a mutated human spastin.
 CC N.B. The present sequence is not given in the present specification but
 CC is derived from the human spastin nucleotide sequence (AAH20174) as
 CC stated on page 14.

XX Sequence 12792 BP; 4163 A; 2256 C; 2487 G; 3886 T; 0 other;

Query Match 100.0%; Score 593; DB 22; Length 12792;
 Best Local Similarity 100.0%; Pred. No. 1.7e-272;

Matches 593; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 acatctatgtttacagagcttctctgtttgtatgaagatagcaacgaaactcaaatggt 60
 DB 12200 acatctatgtttacagagcttctctgtttgtatgaagatagcaacgaaactcaaatggt 12259

OY 61 ggcagttcttattaccaggtttagtattgtttctcgaacactgcttcgaagaacatt 120
 DB 12260 ggcagttcttattaccaggtttagtattgtttctcgaacactgcttcgaagaacatt 12319

OY 121 tattaactgttgaacactgcttattgttgtgtgtacatatttccacaagtgtata 180
 DB 12320 tattaactgttgaacactgcttattgttgtgtgtacatatttccacaagtgtata 12379

OY 181 attatatagttgtgttgaacagatgcaactcttctgttcctaaagtgtcgcagttaa 240
 DB 12380 attatatagttgtgttgaacagatgcaactcttctgttcctaaagtgtcgcagttaa 12439

OY 241 aaaaaaacacacttcttcttcaatctgcatgtagtgagtttcttaacttaaaac 300
 DB 12440 aaaaaaacacacttcttcttcaatctgcatgtagtgagtttcttaacttaaaac 12499

OY 301 atcaaaaatgtttaaatactgtgtatctatagttatgttaattatcgcttatattc 360
 DB 12500 atcaaaaatgtttaaatactgtgtatctatagttatgttaattatcgcttatattc 12559

OY 361 cccatgaatgcatgacactgacattcaattcatgtgtctgcgcacgtcttactt 420
 DB 12560 cccatgaatgcatgacactgacattcaattcatgtgtctgcgcacgtcttactt 12619

OY 421 aacatattcttcttcagaaatgtaaaaggaatgataaattagttatataagtgactcg 480
 DB 12620 aacatattcttcttcagaaatgtaaaaggaatgataaattagttatataagtgactcg 12679

OY 481 ctgtaaatgtgttaatactatttcaatttaagaggttacagaaacatggttgaact 540
 DB 12680 ctgtaaatgtgttaatactatttcaatttaagaggttacagaaacatggttgaact 12739

OY 541 ttttactttatttgggaataagaaatgttgcacactccacatttattgtct 593
 DB 12740 ttttactttatttgggaataagaaatgttgcacactccacatttattgtct 12792

RESULT 3
 AAH20174
 ID AAH20174 standard; DNA; 12793 BP.
 XX
 AC AAH20174;
 XX
 DT 09-AUG-2001 (first entry)
 XX
 DE Human spastin nucleotide sequence SEQ ID NO:1.
 XX
 KW Human; mouse; spastin; ARSACS; chromosome 13q11; identification;
 KW autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;
 KW reduced motor nerve velocity; hypermyelination of retinal nerve fibres;
 KW atrophy of upper cerebellar vermis; absence of Purkinje cell;
 KW abnormal neuronal lipid storage; genetic disorder; characterisation; ds.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 77..11566
 FT /tag= a
 FT /product= "spastin"
 XX
 PN WO200129266-A2.
 XX
 PD 26-APR-2001.
 XX
 PF 20-OCT-2000; 2000WO-US29130.
 XX

PR 20-OCT-1999; 9905-0160588.
 XX
 PA (UYMC-) UNIV MCGILL.
 PA (HOP-) HOPITAL SAINTE-JUSTINE.
 XX
 PI Hudson TJ, Engert J, Richter A;
 XX
 DR WPI: 2001-308494/32.
 XX
 SR P-PSDB; AAB97819.
 XX
 PT New isolated polynucleotide, encoding spastin gene, and polypeptides,
 PT useful for diagnosing autosomal recessive spastic ataxia of
 PT Charlevoix-Saguenay disease by detecting two point mutations in spastin
 PT gene sequence -
 XX
 PS Claim 1; Fig 9; 76pp; English.
 XX
 CC The present invention describes human and mouse spastin, and mutated
 CC human spastin (autosomal recessive spastic ataxia of Charlevoix-Saguenay
 CC (ARSACS)) gene sequences (1). The spastin gene has been mapped to
 CC chromosome 13q11. (1) have neuroprotective activities and can be used in
 CC gene therapy and as a spastin polypeptide agonists. (1), their fragments
 CC or their complements can be useful for assaying the presence of a nucleic
 CC acid molecule in a sample. (1) is useful for diagnosing or aiding in the
 CC diagnosis of an early onset neurodegenerative disease in an individual.
 CC The neurodegenerative disease comprises reduced sensory nerve conduction,
 CC reduced motor nerve velocity, hypermyelination of retinal nerve fibres,
 CC atrophy of upper cerebellar vermis, absence of Purkinje cells and
 CC abnormal neuronal lipid storage. (1) can also be used to produce
 CC antisense nucleic acids, is useful as molecular weight or chromosome
 CC markers, to identify genetic disorders, as hybridisation probes or
 CC primers, as an antigen, identify and express recombinant protein for
 CC analysis, characterisation or therapeutic use, or as markers for tissues
 CC in which the corresponding protein is expressed. Diagnostic methods from
 CC the present invention can be used to identify subjects having or at risk
 CC of developing a disease or disorder associated with aberrant expression
 CC or activity of (1). The assays can be utilised to identify a subject
 CC having or at risk of developing a disorder associated with spastin
 CC protein or spastin gene expression or activity. The present sequence
 CC encodes human spastin as given in the present invention.
 XX
 SQ Sequence 12793 BP; 4163 A; 2256 C; 2487 G; 3887 T; 0 other;

Query Match 100.0%; Score 593; DB 22; Length 12793;
 Best local Similarity 100.0%; Pred. No. 1.7e-272.
 Matches 593; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 acatctatgtttacagagcttctctgtttgtatgaagatagcaacgaaactcaaatggt 60
 DB 12201 acatctatgtttacagagcttctctgtttgtatgaagatagcaacgaaactcaaatggt 12260

OY 61 ggcagttcttattaccaggtttagtattgtttctcgaacactgcttcgaagaacatt 120
 DB 12261 ggcagttcttattaccaggtttagtattgtttctcgaacactgcttcgaagaacatt 12320

OY 121 tattaactgttgaacactgcttattgttgtgtgtacatatttccacaagtgtata 180
 DB 12321 tattaactgttgaacactgcttattgttgtgtgtacatatttccacaagtgtata 12380

OY 181 attatatagttgtgttgaacagatgcaactcttctgttcctaaagtgtcgcagttaa 240
 DB 12381 attatatagttgtgttgaacagatgcaactcttctgttcctaaagtgtcgcagttaa 12440

OY 241 aaaaaaacacacttcttcttcaatctgcatgtagtgagtttcttaacttaaaac 300
 DB 12441 aaaaaaacacacttcttcttcaatctgcatgtagtgagtttcttaacttaaaac 12500

OY 301 atcaaaaatgtttaaatactgtgtatctatagttatgttaattatcgcttatattc 360
 DB 12501 atcaaaaatgtttaaatactgtgtatctatagttatgttaattatcgcttatattc 12560

OY 361 cccatgaatgcatgacactgacattcaattcatgtgtctgcgcacgtcttactt 420

Db 12561 cccatgatgatcagacagcttcaattcatcagttgtctgcgcagcttcttactt 12620
 QY 421 aacatattcttttcgcagaatgtaaaaggaatgataattagttatataagttactgg 480
 Db 12621 aacatattcttttcgcagaatgtaaaaggaatgataattagttatataagttactgg 12680
 QY 481 ctgtaaatgctgctaaatatacttctatgcaattagggtctacagaaatgtgaaact 540
 Db 12681 ctgtaaatgctgctaaatatacttctatgcaattagggtctacagaaatgtgaaact 12740
 QY 541 ttttactttatgtggaataagaatgttgcacccacatttttctt 593
 Db 12741 ttttactttatgtggaataagaatgttgcacccacatttttctt 12793

RESULT 4
 AAH20178
 ID AAH20178 standard; DNA; 12793 BP.
 XX AAH20178;
 AC
 DT 09-AUG-2001 (first entry)
 XX
 DE Human mutated spastin nucleotide sequence SEQ ID NO:11.
 XX
 KW Human; mouse; spastin; ARSACS; chromosome 13q11; identification;
 KW autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;
 KW neurodegenerative disease; reduced sensory nerve conduction; diagnosis;
 KW reduced motor nerve velocity; hypermyelination of retinal nerve fibre;
 KW atrophy of upper cerebellar vermis; absence of Purkinje cell;
 KW abnormal neuronal lipid storage; genetic disorder; characterisation; ds.
 XX
 OS Homo sapiens.
 OS Synthetic.
 XX
 PN MO200129266-A2.
 XX
 PD 26-APR-2001.
 XX
 PF 20-OCT-2000; 2000MO-US29130.
 XX
 PR 20-OCT-1999; 99US-0160588.
 XX
 PA (UYMC-) UNIV MCGILL.
 PA (HOP1-) HOPITAL SAINTE-JUSTINE.
 XX
 PI Hudson TJ, Engert J, Richter A;
 XX
 DR MPI; 2001-308494/32.
 XX
 PT New isolated polynucleotide, encoding spastin gene, and polypeptides,
 PT useful for diagnosing autosomal recessive spastic ataxia of
 PT Charlevoix-Saguenay disease by detecting two point mutations in spastin
 PT gene sequence -
 XX
 PS Claim 1; Page -: 76pp; English.
 XX
 CC The present invention describes human and mouse spastin, and mutated
 CC human spastin (autosomal recessive spastic ataxia of Charlevoix-Saguenay
 CC (ARSACS)) gene sequences (I). The spastin gene has been mapped to
 CC chromosome 13q11. (I) have neuroprotective activities and can be used in
 CC gene therapy and as a spastin polypeptide agonists. (I), their fragments
 CC or their complements can be useful for assaying the presence of a nucleic
 CC acid molecule in a sample. (I) is useful for diagnosing or aiding in the
 CC diagnosis of an early onset neurodegenerative disease in an individual.
 CC The neurodegenerative disease comprises reduced sensory nerve conduction,
 CC reduced motor nerve velocity, hypermyelination of retinal nerve fibres,
 CC atrophy of upper cerebellar vermis, absence of Purkinje cells and
 CC abnormal neuronal lipid storage. (I) can also be used to produce
 CC antisense nucleic acids, is useful as molecular weight or chromosome
 CC markers, to identify genetic disorders, as hybridisation probes or
 CC primers, as an antigen, identify and express recombinant protein for

CC analysis, characterisation or therapeutic use, or as markers for tissues
 CC in which the corresponding protein is expressed. Diagnostic methods from
 CC the present invention can be used to identify subjects having or at risk
 CC of developing a disease or disorder associated with aberrant expression
 CC or activity of (I). The assays can be utilised to identify a subject
 CC having or at risk of developing a disorder associated with spastin
 CC protein or spastin gene expression or activity. The present sequence
 CC represents a mutated human spastin gene from the present invention.
 CC N.B. The present sequence is not given in the present specification but
 CC is derived from the human spastin nucleotide sequence (AAH20174) as
 CC stated on page 14.
 XX
 SQ Sequence 12793 BP; 4163 A; 2255 C; 2487 G; 3888 T; 0 other:

Query Match 100.0%; Score 593; DB 22; Length 12793;
 Best Local Similarity 100.0%; Pred. No. 1.7e-272;
 Matches 593; Conservative 0; Mismatches 0; Indels 0; Caps 0;

QY 1 acatctatgtttacagcttccctgtttgataagatagcaaggaataactcaaatggt 60
 Db 12201 acatctatgtttacagcttccctgtttgataagatagcaaggaataactcaaatggt 12260
 QY 61 ggcagctcttattccagctgttgattgttctcggaaactgttccaaagacaact 120
 Db 12261 ggcagctcttattccagctgttgattgttctcggaaactgttccaaagacaact 12320
 QY 121 tattaactgttagaacacttgcttatagttgtgtgtatataattccacaatgttata 180
 Db 12321 tattaactgttagaacacttgcttatagttgtgtgtatataattccacaatgttata 12380
 QY 181 attatatagtgtgtgtgaacagatgcaactcttgttcttaagtgctgcagttaa 240
 Db 12381 attatatagtgtgtgtgaacagatgcaactcttgttcttaagtgctgcagttaa 12440
 QY 241 aaaaaaaacaaaccttcttccaatgatgagcagctgagtgagtttttaacttaaaac 300
 Db 12441 aaaaaaaacaaaccttcttccaatgatgagcagctgagtgagtttttaacttaaaac 12500
 QY 301 atcaaaaattgttaaaatcaltgtgtatctagtatgttaataatcgcgttatattc 360
 Db 12501 atcaaaaattgttaaaatcaltgtgtatctagtatgttaataatcgcgttatattc 12560
 QY 361 cccatgatgatcagacatgcatltaattcattgttctcgcacagcttcttactt 420
 Db 12561 cccatgatgatcagacatgcatltaattcattgttctcgcacagcttcttactt 12620
 QY 421 aacatattcttttcgcagaatgtaaaaggaatgataattagttatataagttactgg 480
 Db 12621 aacatattcttttcgcagaatgtaaaaggaatgataattagttatataagttactgg 12680
 QY 481 ctgtaaatgctgctaaatatacttctatgcaattagggtctacagaaatgtgaaact 540
 Db 12681 ctgtaaatgctgctaaatatacttctatgcaattagggtctacagaaatgtgaaact 12740
 QY 541 ttttactttatgtggaataagaatgttgcacccacatttttctt 593
 Db 12741 ttttactttatgtggaataagaatgttgcacccacatttttctt 12793

RESULT 5
 AAH20179
 ID AAH20179 standard; DNA; 12793 BP.
 XX AAH20179;
 AC
 DT 09-AUG-2001 (first entry)
 XX
 DE Human mutated spastin nucleotide sequence SEQ ID NO:12.
 XX
 KW Human; mouse; spastin; ARSACS; chromosome 13q11; identification;
 KW autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;
 KW neurodegenerative disease; reduced sensory nerve conduction; diagnosis;

XX	reduced motor nerve velocity; hypermyelination of retinal nerve fibre;
KW	atrophy of upper cerebellar vermis; absence of Purkinje cell;
KW	abnormal neuronal lipid storage; genetic disorder; characterisation; ds.
XX	
OS	Homo sapiens.
OS	Synthetic.
XX	
PN	WO200129266-A2.
XX	
PD	26-APR-2001.
XX	
PE	20-OCT-2000; 2000MO-US29130.
XX	
PR	20-OCT-1999; 99US-0160588.
XX	
XX	
PA	(UYMC-) UNIV MCGILL.
PA	(HOPIT-) HOPITAL SAINTE-JUSTINE.
XX	
PI	Hudson TJ, Engert J, Richter A;
XX	
DR	WPI; 2001-308494/32.
XX	
PT	New isolated polynucleotide, encoding spastin gene, and polypeptides,
PT	useful for diagnosing autosomal recessive spastic ataxia of
PT	Charlevoix-Saguenay disease by detecting two point mutations in spastin
PT	gene sequence -
XX	
OS	Claim 1; Page -: 76pp; English.

The present invention describes human and mouse spastin, and mutated human spastin (autosomal recessive spastic ataxia of Charlevoix-Saguenay (ARSACS)) gene sequences (1). The spastin gene has been mapped to chromosome 13q11. (1) have neuroprotective activities and can be used in gene therapy and as a spastin polypeptide agonists. (1), their fragments or their complements can be useful for assaying the presence of a nucleic acid molecule in a sample. (1) is useful for diagnosing or aiding in the diagnosis of an early onset neurodegenerative disease in an individual. The neurodegenerative disease comprises reduced sensory nerve conduction, reduced motor nerve velocity, hypomyelination of retinal nerve fibres, atrophy of upper cerebellar vermis, absence of Purkinje cells and abnormal neuronal lipid storage. (1) can also be used to produce antisense nucleic acids, is useful as molecular weight or chromosome markers, to identify genetic disorders, as hybridisation probes or primers, as an antigen, identify and express recombinant protein for analysis, characterisation or therapeutic use, or as markers for tissues in which the corresponding protein is expressed. Diagnostic methods from the present invention can be used to identify subjects having or at risk of developing a disease or disorder associated with aberrant expression or activity of (1). The assays can be utilised to identify a subject having or at risk of developing a disorder associated with spastin protein or spastin gene expression or activity. The present sequence encodes a mutated human spastin.

N.B. The present sequence is not given in the present specification but is derived from the human spastin nucleotide sequence (AM20174) as stated on page 14.

Sequence 12793 BP; 4163 A; 2257 C; 2487 G; 3886 T; 0 other;

Query Match	100.0%	Score 593;	DB 22;	Length 12793;
Best Local Similarity	100.0%	Pred. NO. 1.7e-272;		
Matches 593;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY	1	acctcttatgtttcaaggcttcctctgtttgtaagaatagcacaaggaaaacctcaaatgct	60
Db	12201	acacctcttgtttcaaggcttcctctgtttgtaagaatagcacaaggaaaacctcaaatgct	12260
QY	61	gcacgtttcttataccagctgttgaagtatgtttctgcgaacacgcttcgcgaagacacatt	120
Db	12261	gcacgtttcttataccagctgttgaagtatgtttctgcgaacacgcttcgcgaagacacatt	12320
QY	121	tatttaacgttagaacaacttgcttaagtctgtgtgtacataattctccaaaatgttata	180
Db	1221	tatttaacgttagaacaacttgcttaagtctgtgtgtacataattctccaaaatgttata	180

Db	12321	tataacggttagaacacttgctttaigtgttggtgtagacataattccacaatgltaba	12380
Qy	181	attatatacagtgvtgvtgtaacacagatcgtaacctttgtgtctctaaagvtgcgtcagttaa	240
Db	12381	atttatatacgtgtgtgtaacacagatcgtaacctttgtgtctctaaagvtgcgtcagttaa	12440
Qy	241	aaaaaaaaaacacctttcttccaatagcgtgtgtgtgttttttaactttaaanaac	300
Db	12441	aaaaaaaaaacacctttcttccaatagcgtgtgtgtgttttttaactttaaanaac	12500
Qy	301	atcaaaaatctgttaaaccatcgtgtgttactagtagttacaataatcgcgtcaatttc	360
Db	12501	atcaaaaatctgttaaaccatcgtgtgttactagtagttacaataatcgcgtcaatttc	12560
Qy	361	cccatgtaatgatcagaaactgacatttaatactgtttgtccgcgaatgctcttaacttc	420
Db	12561	cccatgtaatgatcagaaactgacatttaatactgtttgtccgcgaatgctcttaacttc	12620
Qy	421	aacatattctcttgcagaaatgtaaaagtgatgttaagtattataaagttaactg	480
Db	12621	aacatattctcttgcagaaatgtaaaagtgatgttaagtattataaagttaactg	12680
Qy	481	ctgttaaatgagtctaaataatacttatactgtcaattaaaggcttaacagaaatgttgaactt	540
Db	12681	ctgttaaatgagtctaaataatacttatactgtcaattaaaggcttaacagaaatgttgaactt	12740
Qy	541	tttttaactttatgtggaataagaatgtttggaactccaatttatgtct	593
Db	12741	tttttaactttatgtggaataagaatgtttggaactccaatttatgtct	12793

XX	RESULT	6
XX	AAH20182	
ID	AAH20182 standard; DNA; 12793 BP.	
XX	AAH20182;	
XX	09-AUG-2001 (first entry)	
DE	Human mutated spastin nucleotide sequence SEQ ID NO:15.	
XX		
XX	Human; mouse; spastin; ARSACS; chromosome 13q11; identification;	
KW	autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;	
KW	neurodegenerative disease; reduced sensory nerve conduction; diagnosis;	
KW	reduced motor nerve velocity; hypermyelination of retinal nerve fibre;	
KW	atrophy of upper cerebellar vermis; absence of Purkinje cell;	
KW	abnormal neuronal lipid storage; genetic disorder; characterisation; ds	
XX		
OS	Homo sapiens.	
OS	Synthetic.	
XX		
XX	Key	Location/Qualifiers
FT	CDS	77..11566
FT		/*tag= a
FT		/product= "mutated spastin"
XX		
XX	MO200129266-A2.	
XX		
XX	26-APR-2001.	
XX		
XX	20-OCT-2000; 2000MO-US29130.	
XX		
XX	20-OCT-1999; 99US-0160588.	
XX		
XX	(UYMC-) UNIV MCGILL.	
XX	(HOP1-) HOPITAL SAINTE-JUSTINE.	
XX		
PI	Hudson TJ, Engert J, Richter A;	
XX		
XX	WPI: 2001-308494/32.	
DR	P-PSDB: AAB97823.	
XX		
XX	New isolated polynucleotide, encoding spastin gene, and polypeptides,	

PT useful for diagnosing autosomal recessive spastic ataxia of
PT Charlevoix-Saguenay disease by detecting two point mutations in spastin
PT gene sequence -
PS
PS Claim 1: Page -: 76pp: English.
XX
XX The present invention describes human and mouse spastin, and mutated
CC human spastin (autosomal recessive spastic ataxia of Charlevoix-Saguenay
CC (ARSACS)) gene sequences (I). The spastin gene has been mapped to
CC chromosome 13q11. (I) have neuroprotective activities and can be used in
CC gene therapy and as a spastin polypeptide agonists. (I), their fragments
CC or their complements can be useful for assaying the presence of a nucleic
CC acid molecule in a sample. (I) is useful for diagnosing or aiding in the
CC diagnosis of an early onset neurodegenerative disease in an individual.
CC The neurodegenerative disease comprises reduced sensory nerve conduction,
CC reduced motor nerve velocity, hypermyelination of retinal nerve fibres,
CC atrophy of upper cerebellar vermis, absence of Purkinje cells and
CC abnormal neuronal lipid storage. (I) can also be used to produce
CC antisense nucleic acids, is useful as molecular weight or chromosome
CC markers, to identify genetic disorders, as hybridisation probes or
CC primers, as an antigen, identify and express recombinant protein for
CC analysis, characterisation or therapeutic use, or as markers for tissues
CC in which the corresponding protein is expressed. Diagnostic methods from
CC the present invention can be used to identify subjects having or at risk
CC of developing a disease or disorder associated with aberrant expression
CC or activity of (I). The assays can be utilised to identify a subject
CC having or at risk of developing a disorder associated with Spastin
CC protein or spastin gene expression or activity. The present sequence
CC encodes a mutated human spastin.
CC N.B. The present sequence is not given in the present specification but
CC is derived from the human spastin nucleotide sequence (AAH20174) as
CC stated on page 14.
XX
XX Sequence 12793 BP, 4163 A; 2257 C; 2487 G; 3886 T; 0 other:

Query Match 100.0%; Score 593; DB 22; Length 12793;
Best Local Similarity 100.0%; Pred. No. 1.7e-272;
Matches 593; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 acatctatggttaccaggtctctgttgaagaagatgacaagaaactcaaatggt 60
DB 12201 acatctatggttaccaggtctctgttgaagaagatgacaagaaactcaaatggt 12260
QY 61 ggcagttcttattaccaggttgaatgattgttcttgcgaactctgcgaagaacatt 120
DB 12261 ggcagttcttattaccaggttgaatgattgttcttgcgaactctgcgaagaacatt 12320
QY 121 tattaactgttagaacaactgcttattgttgggttacctatttccacaatgttata 180
DB 12321 tattaactgttagaacaactgcttattgttgggttacctatttccacaatgttata 12380
QY 181 atttataatggttgaacagatgacaacttctgttgcacaaagttgcagttaa 240
DB 12381 atttataatggttgaacagatgacaacttctgttgcacaaagttgcagttaa 12440
QY 241 aaaaaaaacaaccttcttcaatatgacatgtagtgcagtttctttaaactaaac 300
DB 12441 aaaaaaaacaaccttcttcaatatgacatgtagtgcagtttctttaaactaaac 12500
QY 301 atcaaaatgtttaaatatctgttattatgatttattttaaattgcgttataattc 360
DB 12501 atcaaaatgtttaaatatctgttattatgatttattttaaattgcgttataattc 12560
QY 361 cccatgaatgatcagaactgacattcaatcagttgttcgcacatgctcttcaatt 420
DB 12561 cccatgaatgatcagaactgacattcaatcagttgttcgcacatgctcttcaatt 12620
QY 421 aacataatcttcttgagaagtgttaaaagtgaaatgaataatgattatataagtgctacg 480
DB 12621 aacataatcttcttgagaagtgttaaaagtgaaatgaataatgattatataagtgctacg 12680
QY 481 ctgtaaatgatgctaataacttattgcaattaaggcttacaagaacatgttgaactt 540

DB 12681 ctgtaaatgatgctaataacttattgcaattaaggcttacaagaacatgttgaactt 12740
QY 541 ttttactttattgggaataaggaatgtttgcacccatcatattatgctt 593
DB 12741 ttttactttattgggaataaggaatgtttgcacccatcatattatgctt 12793
RESULT 7
ID AAS29132 standard; CDNA; 1387 BP.
XX
XX AAS29132:
AC
XX 21-NOV-2001 (first entry)
DT
XX
XX cDNA encoding for human DNA-binding protein #103.
DE
XX
XX Human, DNA-binding protein; histone; chromo domain protein;
KW chromatin organisation modifier; r-box binding protein;
KW DNA organisation; gene transcription; malignant disease;
KW autoimmune disorder; rheumatic disease; genetic abnormality;
KW infectious disease; neurological disorder; gene therapy;
KW immunomodulatory; anti-HIV; anti rheumatic; anti microbial;
KW cytostatic; ss.
XX
XX Homo sapiens.
OS
XX
XX WO200155162-A1.
PN
XX
XX 02-AUG-2001.
PD
XX
XX 17-JAN-2001; 2001MO-US01305.
PE
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.

```

PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239335.
PR 13-OCT-2000; 2000US-0239337.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.

```

```

PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249219.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249266.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PI Rosen CA, Barash SC, Ruben SM.
XX
XX WPI; 2001-465557/50.
DR P-PSDB; AAU18256.
DR
XX
PT Nucleic acid molecules encoding human secreted chromosomal binding
PT proteins, used in preventing, treating or ameliorating a disorder, e.g.
PT Alzheimer's and Parkinson's diseases and cancers -
PT
XX
XX Claim 4; SEQ ID NO 113; 561pp; English.
XX
CC The present invention relates to the isolation of novel DNA-binding
CC proteins (AAU18154-AAU18281), and cDNA and genomic sequences encoding
CC for these proteins. DNA-binding proteins such as histones, chromo
CC (chromatin organisation modifier) domain proteins, and Y-box binding
CC proteins may contribute to diseases resulting from aberrant DNA
CC organisation and/or gene transcription. The sequences of the invention
CC are useful in screening assays to identify antagonists and/or agonists
CC that may enhance or block activities mediated by DNA-binding proteins.
CC Blockers of DNA-binding proteins may be useful in treating disorders
CC such as malignant diseases (e.g. cancer), autoimmune disorders
CC (e.g. diabetes mellitus), rheumatic diseases (e.g. rheumatoid
CC arthritis), genetic abnormalities (e.g. cystic fibrosis), infectious
CC diseases (e.g. HIV) and neurological disorders (e.g. Alzheimer's
CC disease). The polynucleotide sequences of the invention may also be
CC used in gene therapy. MAS29030-MAS29157 represent cDNA sequences
CC encoding for novel DNA-binding proteins.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1387 BP; 494 A; 165 C; 238 G; 482 T; 8 other;

```

```

Query Match 72.5%; Score 430; DB 22; Length 1387;
Best Local Similarity 100.0%; Pred. No. 7.1e-195;
Matches 430; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 164 ttccacaatgtataattatagtggtggaacagatgacatcttggctc 223
DB 782 ttccacaatgtataattatagtggtggaacagatgacatcttggctc 841
QY 224 aaagtgctgcagtttaaaaaaacacctttcttcaatgtgcatgtgagtt 283
DB 842 aaagtgctgcagtttaaaaaaacacctttcttcaatgtgcatgtgagtt 901

```


Qy	284	ttttaactttaaaacatcaaaaaattgtttaaaatcatgtgttacttcagtagttataa	343
Db	902	ttttaactttaaaacatcaaaaaattgtttaaaatcatgtgttacttcagtagttataa	961
Qy	344	ttatcgcttatatttcccatgaatgacacgaactcgaatttaattcatgtttgtctcg	403
Db	962	ttatcgcttatatttcccatgaatgacacgaactcgaatttaattcatgtttgtctcg	1021
Qy	404	ccatgcctcttaactttaacatattctcttcgcagatgttaaagtaagtaataatagt	463
Db	1022	ccatgcctctcttaactttaacatattctcttcgcagatgttaaagtaagtaataatagt	1081
Qy	464	ttatataagtgtaactgctgctgtaaatgatactcaataatacttatctcaattaaagggcttac	523
Db	1082	ttatataagtgtaactgctgctgtaaatgatactcaataatacttatctcaattaaagggcttac	1141
Qy	524	agaacatgcttgaaacttttttaacttttaattggygaataaggaagtcttcgacctccacat	583
Db	1142	agaacatgcttgaaacttttttaacttttaattggygaataaggaagtcttcgacctccacat	1201
Qy	584	tttatcgctt	593
Db	1202	tttatcgctt	1211

RESULT	8
AAI76962	
ID	AAI76962 standard; DNA; 51 BP.
XX	
AC	AAI76962;
XX	
DT	09-NOV-2001 (first entry)
XX	
DE	Human silent SNP containing nucleic acid SEQ:3903.
XX	
KW	Human; single nucleotide polymorphism; SNP; genome; gene therapy;
KM	protein therapy; vaccine; probe; diagnostic assay; detection;
XX	quantitation; restorative therapy; polymorphic; ds.
OS	Homo sapiens.
XX	
PN	WO200140521-A2.
PD	
XX	07-JUN-2001.
PF	
XX	30-NOV-2000; 2000MO-US32758.
PR	30-NOV-1999; 99US-0168138.
XX	29-NOV-2000; 2000US-0726173.
PA	(CURA-) CURAGEN CORP.
XX	
P1	Shinkets RA, Leach M;
DR	WPI; 2001-356160/37.
XX	
PT	Polymorphic nucleic acid sequences, useful in genetic testing and
FT	therapy -
PS	
XX	Claim 1; Page 1246; 2653pp; English.
CC	AAI79862 represent isolated human polymorphic polynucleotide
CC	sequences (I), which contain single nucleotide polymorphisms (SNPs).
CC	AAM5314 to AAM5339 represent peptides related to human polymorphic
CC	polynucleotide sequences. The sequences can be used in gene and protein
CC	therapy, and in vaccine production. (I) and the polypeptides encoded by
CC	them may be used in the prevention, diagnosis and treatment of diseases
CC	associated with inappropriate expression of polymorphic polypeptides.
CC	For example, (I) may be used to treat disorders by rectifying mutations
CC	or deletions in a patient's genome that affect the activity of
CC	polypeptides by expressing inactive proteins or to supplement the
CC	patients own production of polypeptide. Additionally, (I) and its
CC	complementary sequences may also be used as DNA probes in diagnostic

assays to detect and quantitate the presence of similar nucleic acids in samples, and therefore which patients may be in need of restorative therapy. The polypeptides encoded by (1) may be used as antigens in the production of antibodies specific for polymorphic polypeptides. The antibodies may also be used to down regulate expression and activity. The antibodies may also be used as diagnostic agents for detecting the presence of polymorphic polypeptides in samples.

Query Match	8.6%	Score 51;	DB 22;	Length 51;
Best Local Similarity	100.0%	Pred. No. 1.9e-14;		
Matches	51;	Conservative	0;	Mismatches 0;
			Indels	0;
			Gaps	0;

QY 109 caagacaacatttattactgttagaacactgcttatgtgtgtgtac 159
|||||
Db 1 caagacaacatttattactgttagaacactgcttatgtgtgtgtac 51

RESULT	9
AAI76964	
ID	AAI76964 standard; DNA; 51 BP.

09-NOV-2001 (first entry)
Human silent SNP containing nucleic acid SEQ:3905.
Human: single nucleotide polymorphism; SNP; genome; gene therapy;
protein therapy; vaccine; probe; diagnostic assay; detection;
quantitation; restorative therapy; polymorphic; ds.
Homo sapiens.
WO200140521-A2.
07-JUN-2001.
30-NOV-2000; 2000WO-US32758.
30-NOV-1999; 99US-0168138.
29-NOV-2000; 2000US-0726173.
(CURA-) CURAGEN CORP.
Shimkets RA, Leach M;
WPI: 2001-356160/37.
Polymorphic nucleic acid sequences, useful in genetic testing and
therapy -
Claim 1, Page 1246; 2653pp; English.

CC The antibodies may also be used as diagnostic agents for detecting the
CC presence of polymorphic polypeptides in samples.
XX
SQ Sequence 51 BP; 20 A; 8 C; 8 G; 15 T; 0 other;

Query Match 8.6%; Score 51; DB 22; Length 51;
Best Local Similarity 100.0%; Pred. No. 1.9e-14;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 223 taagggtcgtcagcttaaaaaaacaccccttcttccaatagcagc 273
Db 1 taagggtcgtcagcttaaaaaaacaccccttcttccaatagcagc 51

RESULT 10

AAI76965
ID AAI76965 standard; DNA; 50 BP.

AC AAI76965;

XX 09-NOV-2001 (first entry)

DE Human silent SNP containing nucleic acid SEQ:3906.

XX Human; single nucleotide polymorphism; SNP; genome; gene therapy;

KW protein therapy; vaccine; probe; diagnostic assay; detection;

KM quantitation; restorative therapy; polymorphic; ds.

XX Homo sapiens.

XX WO200140521-A2.

XX 07-JUN-2001.

XX 30-NOV-2000; 2000WO-US32758.

XX 30-NOV-1999; 99US-0168138.

PR 29-NOV-2000; 2000US-0726173.

XX (CURA-) CURAGEN CORP.

PA Shinkets RA, Leach M;

PI Shinkets RA, Leach M;

XX WPI; 2001-356160/37.

PT Polymorphic nucleic acid sequences, useful in genetic testing and
therapy -

PS Claim 1; Page 1247; 2653pp; English.

XX AAI73060 to AAI79867 represent isolated human polymorphic polynucleotide
sequences (I), which contain single nucleotide polymorphisms (SNPs).

CC AAM53114 to AAM53329 represent peptides related to human polymorphic
polynucleotide sequences. The sequences can be used in gene and protein
therapy, and in vaccine production. (I) and the polypeptides encoded by
CC them may be used in the prevention, diagnosis and treatment of diseases
associated with inappropriate expression of polymorphic polypeptides.

CC For example, (I) may be used to treat disorders by rectifying mutations
or deletions in a patient's genome that affect the activity of
CC polypeptides by expressing inactive proteins or to supplement the
CC patients own production of polypeptide. Additionally, (I) and its
CC complementary sequences may also be used as DNA probes in diagnostic
assays to detect and quantitate the presence of similar nucleic acids
CC in samples, and therefore which patients may be in need of restorative
therapy. The polypeptides encoded by (I) may be used as antigens in the
CC production of antibodies specific for polymorphic polypeptides. The
CC antibodies may also be used to down regulate expression and activity.
CC The antibodies may also be used as diagnostic agents for detecting the
presence of polymorphic polypeptides in samples.

XX Sequence 50 BP; 19 A; 8 C; 8 G; 15 T; 0 other;

Query Match 5.7%; Score 34; DB 22; Length 50;
Best Local Similarity 100.0%; Pred. No. 2.4e-06;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 240 aaaaaaacaccccttcttccaatagcagc 273
Db 17 aaaaaaacaccccttcttccaatagcagc 50

RESULT 11

AAI76963
ID AAI76963 standard; DNA; 51 BP.

XX AAI76963;

XX 09-NOV-2001 (first entry)

DE Human silent SNP containing nucleic acid SEQ:3904.

XX Human; single nucleotide polymorphism; SNP; genome; gene therapy;

KW protein therapy; vaccine; probe; diagnostic assay; detection;

KM quantitation; restorative therapy; polymorphic; ds.

XX Homo sapiens.

XX WO200140521-A2.

XX 07-JUN-2001.

XX 30-NOV-2000; 2000WO-US32758.

XX 30-NOV-1999; 99US-0168138.

PR 29-NOV-2000; 2000US-0726173.

XX (CURA-) CURAGEN CORP.

PA Shinkets RA, Leach M;

PI Shinkets RA, Leach M;

XX WPI; 2001-356160/37.

PT Polymorphic nucleic acid sequences, useful in genetic testing and
therapy -

PS Claim 1; Page 1246; 2653pp; English.

XX AAI73060 to AAI79867 represent isolated human polymorphic polynucleotide
sequences (I), which contain single nucleotide polymorphisms (SNPs).

CC AAM53114 to AAM53329 represent peptides related to human polymorphic
polynucleotide sequences. The sequences can be used in gene and protein
therapy, and in vaccine production. (I) and the polypeptides encoded by
CC them may be used in the prevention, diagnosis and treatment of diseases
associated with inappropriate expression of polymorphic polypeptides.

CC For example, (I) may be used to treat disorders by rectifying mutations
or deletions in a patient's genome that affect the activity of
CC polypeptides by expressing inactive proteins or to supplement the
CC patients own production of polypeptide. Additionally, (I) and its
CC complementary sequences may also be used as DNA probes in diagnostic
assays to detect and quantitate the presence of similar nucleic acids
CC in samples, and therefore which patients may be in need of restorative
therapy. The polypeptides encoded by (I) may be used as antigens in the
CC production of antibodies specific for polymorphic polypeptides. The
CC antibodies may also be used to down regulate expression and activity.
CC The antibodies may also be used as diagnostic agents for detecting the
presence of polymorphic polypeptides in samples.

XX Sequence 51 BP; 14 A; 8 C; 9 G; 20 T; 0 other;

Query Match 4.2%; Score 25; DB 22; Length 51;
Best Local Similarity 100.0%; Pred. No. 0.045;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;


```

XX 18-DEC-2001 (First entry)
DE Chemically pretreated complementary DNA associated with cell cycle #98.
XX
XX
XX Cell cycle; human; Cpg dinucleotide; cytosine methylation; HIV; aging;
XX human immunodeficiency virus; neurodegenerative disorder; solid tumour;
XX graft-versus-host disease; glomerular disease; Lewy body disease; cancer;
XX arthritis; arteriosclerosis; anti-HIV; neuroprotective; antiarthritic;
XX immunosuppressive; antitumour; cytostatic; antiarteriosclerotic; ds;
XX PCR primer.
XX
XX Homo sapiens.
XX
XX WO200168911-A2.
XX
XX 20-SEP-2001.
XX
XX 15-MAR-2001; 2001WO-EP02945.
XX
XX 15-MAR-2000; 2000DE-1013847.
XX 06-APR-2000; 2000DE-1019058.
XX 07-APR-2000; 2000DE-1019173.
XX 30-JUN-2000; 2000DE-1032529.
XX 01-SEP-2000; 2000DE-1043826.
XX
XX (EPIC-) EPIGENOMICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K;
XX
XX WPI; 2001-602751/68.
XX
XX Designing primers and probes for analysing diseases associated with
XX cytosine methylation state e.g. arthritis; cancer, aging
XX arteriosclerosis comprising fragments of chemically modified genes
XX associated with cell cycle -
XX
XX Claim 1; SEQ ID No 196; 28pp; English.
XX
XX Sequences AAS45296-AAS45520 represent chemically pretreated genomic DNA
XX molecules associated with the cell cycle and specific PCR primers of the
XX invention. The sequences are useful for detecting the methylation state
XX of all CpG dinucleotides in a sequence and therefore for analysing
XX associated diseases. By analysing cytosine methylations in the pretreated
XX DNA, genetic and/or epigenetic parameters for the diagnosis and therapy
XX of existing diseases or the predisposition to specific diseases can be
XX ascertained. The parameters may be compared to another set of genetic
XX and/or epigenetic parameters, the differences serving as basis for
XX diagnosis and/or prognosis events which are disadvantageous to patients.
XX The sequences of the invention are useful for the diagnosis and therapy
XX of HIV infection, neurodegenerative disorders, graft-versus-host disease,
XX aging, glomerular disease, Lewy body disease, arthritis,
XX arteriosclerosis, solid tumours and cancers.
XX
XX Sequence 4604 BP; 1054 A; 181 C; 1144 G; 2225 T; 0 other;
SQ

```

```

Query Match 3.5%; Score 21; DB 22; Length 4604;
Best Local Similarity 100.0%; Pred. No. 3.1;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 242 aaaaacacaccttcttc 262
   ||||||||||||||||
Db 2569 AAAAAAACACCTTTCTTC 2549

```

```

RESULT 15
AAV90383
ID AAV90383 standard; cDNA; 268 BP.
XX
AC AAV90383;
XX
DT 15-FEB-1999 (first entry)

```

```

XX EST clone DM420.
DE
XX
XX Human; secreted protein; expressed sequence tag; EST; haematopoiesis;
XX tissue growth; activin; inhibin; chemokinesis; chemostatic;
XX receptor; ligand; thrombolytic; anti-inflammatory; cadherin; anti-tumour;
XX gene therapy; ss.
XX
XX Homo sapiens.
XX
XX WO9845436-A2.
XX
XX 15-OCT-1998.
XX
XX 10-APR-1998; 98WO-US06955.
XX
XX 10-APR-1997; 97US-0838821.
XX
XX (GEMY ) GENETICS INST INC.
XX
XX Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;
XX Racie LA, Spaulding V, Treacy M;
XX
XX WPI; 1999-070077/06.
XX
XX New polynucleotides encoding human secreted proteins - derived from
XX e.g. human blood, kidney, foetal lung, placenta, testes, brain,
XX ovary, pituitary, retina and colon cDNA libraries.
XX
XX Claim 1; Page 523; 618pp; English.
XX
XX The present sequence represents a human expressed sequence tag (EST).
XX The polynucleotide, which is a secreted EST, and the encoded protein
XX are predicted to have useful biological activities which would make
XX them suitable for treating, preventing or ameliorating medical
XX conditions in humans and animals, although no supporting data is
XX given. Suggested activities include nutritional activity, immune
XX stimulating or suppressing activity, haematopoiesis regulating
XX activity, tissue growth activity, activin/inhibin activity,
XX chemotactic/chemokinetic activity, haemostatic and thrombolytic
XX activity, receptor/ligand activity, anti-inflammatory activity,
XX cadherin/tumour invasion suppressor activity, tumour inhibition
XX activity. The polynucleotide may also be useful for gene therapy.
XX
XX Sequence 268 BP; 92 A; 47 C; 45 G; 84 T; 0 other;
SQ

```

```

Query Match 3.4%; Score 20; DB 20; Length 268;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 235 agttaaaaaaaaaaacact 254
   ||||||||||||||||
Db 176 agttaaaaaaaaaaacact 195

```


THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 22, 2002, 08:31:08 ; Search time 91.58 Seconds
(without alignments)
1590.529 Million cell updates/sec

Title: US-09-693-205-7_COPY_12200_12792

Sequence: 1 acattcttggttaacagctc.....acctccacatttattgctt 593

Scoring table:

OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 383533 seqs, 122816752 residues

Word size : 0

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

Issued Patents NA: *
1: /cgn2_6/prodata/1/lna/5A.COMB.seq: *
2: /cgn2_6/prodata/1/lna/5B.COMB.seq: *
3: /cgn2_6/prodata/1/lna/5A.COMB.seq: *
4: /cgn2_6/prodata/1/lna/5B.COMB.seq: *
5: /cgn2_6/prodata/1/lna/5A.COMB.seq: *
6: /cgn2_6/prodata/1/lna/5B.COMB.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	18	3.0	6344	4 US-08-843-417-1	Sequence 1, Appl
C 2	17	2.9	164	4 US-09-172-711-56	Sequence 56, Appl
C 3	17	2.9	634	1 US-08-450-065-1	Sequence 1, Appl
C 4	17	2.9	634	1 US-08-450-595-1	Sequence 1, Appl
C 5	17	2.9	2659	1 US-08-007-775-3	Sequence 3, Appl
C 6	17	2.9	2848	2 US-08-805-918-1	Sequence 1, Appl
C 7	17	2.9	3417	2 US-08-464-402-1	Sequence 1, Appl
C 8	17	2.9	3417	4 US-09-054-775C-1	Sequence 1, Appl
C 9	17	2.9	4106	2 US-08-702-572-14	Sequence 14, Appl
C 10	17	2.9	4732	6 5521093-4	Patent No. 5521093
C 11	17	2.9	6038	4 US-09-305-639-4	Sequence 4, Appl
C 12	17	2.9	7622	4 US-09-305-639-1	Sequence 1, Appl
C 13	16	2.7	29	3 US-08-835-728D-96	Sequence 96, Appl
C 14	16	2.7	29	3 US-08-835-728D-200	Sequence 200, App
C 15	16	2.7	29	4 US-09-490-558-96	Sequence 96, App
C 16	16	2.7	29	4 US-09-490-558-200	Sequence 200, App
C 17	16	2.7	285	3 US-09-284-782-34	Sequence 34, Appl
C 18	16	2.7	285	3 US-09-284-782-35	Sequence 35, Appl
C 19	16	2.7	313	1 US-08-365-981-6	Sequence 6, Appl
C 20	16	2.7	633	1 US-08-234-783-1	Sequence 1, Appl
C 21	16	2.7	633	1 US-08-456-907-1	Sequence 1, Appl
C 22	16	2.7	633	5 PCT-US95-05523-1	Sequence 1, Appl
C 23	16	2.7	633	4 US-09-385-982-90	Sequence 90, Appl
C 24	16	2.7	675	1 US-08-307-499-55	Sequence 55, Appl
C 25	16	2.7	675	4 US-09-299-268-55	Sequence 55, Appl
C 26	16	2.7	856	1 US-08-117-373-10	Sequence 10, Appl
C 27	16	2.7	1065	3 US-08-591-685-6	Sequence 6, Appl

C 28	16	2.7	1288	1 US-08-047-041A-24	Sequence 24, Appl
C 29	16	2.7	1313	1 US-08-446-925-6	Sequence 6, Appl
C 30	16	2.7	1313	2 US-09-146-331-6	Sequence 6, Appl
C 31	16	2.7	1313	2 US-08-896-885-6	Sequence 6, Appl
C 32	16	2.7	1313	4 US-09-375-256-6	Sequence 6, Appl
C 33	16	2.7	1316	1 US-08-047-041A-11	Sequence 11, Appl
C 34	16	2.7	1316	2 US-08-795-006A-31	Sequence 31, Appl
C 35	16	2.7	1316	4 US-09-184-073-31	Sequence 31, Appl
C 36	16	2.7	1357	3 US-09-043-123-3	Sequence 3, Appl
C 37	16	2.7	1557	5 PCT-US96-05800-6	Sequence 6, Appl
C 38	16	2.7	1670	3 US-09-026-482B-1	Sequence 1, Appl
C 39	16	2.7	1684	2 US-08-899-811-21	Sequence 21, Appl
C 40	16	2.7	2504	1 US-08-484-105-15	Sequence 15, Appl
C 41	16	2.7	2504	1 US-08-484-106-15	Sequence 15, Appl
C 42	16	2.7	3480	1 US-07-657-769B-68	Sequence 68, Appl
C 43	16	2.7	3480	1 US-07-789-184-219	Sequence 219, App
C 44	16	2.7	3480	1 US-08-475-263-219	Sequence 219, App
C 45	16	2.7	3480	1 US-08-485-086-219	Sequence 219, App

ALIGNMENTS

```
RESULT 1
US-08-843-417-1/C
Sequence 1, Application US/08843417
Patent No. 6184349
GENERAL INFORMATION:
APPLICANT: Herman, Ronald C
APPLICANT: Delgado, Stephen G
APPLICANT: Fish, Linda M
APPLICANT: Sangameswaran, Lakshmi
APPLICANT: Rabert, Douglas K
TITLE OF INVENTION: CLONED PERIPHERAL NERVE
TITLE OF INVENTION: TETRODOXIN-RESISTANT SODIUM CHANNEL alpha-SUBUNIT
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESSES:
ADDRESS: Helier Ehirman White & Mcauliffe
STREET: 525 University Ave
CITY: Palo Alto
STATE: CA
COUNTRY: U.S.A.
ZIP: 94301
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/843,417
FILING DATE: April 15, 1997
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Schmoossee, William
REGISTRATION NUMBER: 31,796
REFERENCE/DOCKET NUMBER: 28340-P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415)-324-7041
TELEFAX: (415)-324-0638
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 6344 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: rat
TISSUE TYPE: Dorsal root ganglia
CELL TYPE: Peripheral nerve
US-08-843-417-1
```

Query Match 3.0%; Score 18; DB 4; Length 6344;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 282 ttttttaacttaaaaa 299
Db 6332 TTTTAACTTAAAAA 6315

RESULT 2
US-09-172-711-56
; Sequence 56, Application US/09172711
; Patent No. 6160105
; GENERAL INFORMATION:
; APPLICANT: Cunningham, Mary Jane
; APPLICANT: Zweiger, Gary B.
; APPLICANT: Panzer, Scott R.
; APPLICANT: Seilhamer, Jeffrey J.
; TITLE OF INVENTION: MONITORING TOXICOLOGICAL RESPONSES
; FILE REFERENCE: PA-0011 US
; CURRENT APPLICATION NUMBER: US/09/172,711
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PERL Program
; SEQ ID NO 56
; LENGTH: 164
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE: -
; OTHER INFORMATION: 700625315H1
US-09-172-711-56

Query Match 2.9%; Score 17; DB 4; Length 164;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 538 ctttttactttatt 554
Db 140 ctttttactttatt 156

RESULT 3
US-08-450-065-1/C
; Sequence 1, Application US/08450065
; Patent No. 5798105
; GENERAL INFORMATION:
; APPLICANT: Schoenmakers, Johannes G
; APPLICANT: Koniings, Rudolph NH
; APPLICANT: Moelans, Inge IMD
; TITLE OF INVENTION: No. 5798105el protein
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Smithkline Beecham Corporate Patents -US
; STREET: UW2220, Po Box 1539
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/450,065
; FILING DATE: 25-MAY-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/949645
; FILING DATE: 04-DEC-1992

ATTORNEY/AGENT INFORMATION:
; NAME: Jervis, Herbert H.
; REGISTRATION NUMBER: 31,171
; REFERENCE/DOCKET NUMBER: B2992
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-270-5065
; TELEFAX: 215-270-5090

; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 634 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: PLASMODIUM
; STRAIN: FALCIPARUM
; IMMEDIATE SOURCE:
; CLONE: 16K
US-08-450-065-1

Query Match 2.9%; Score 17; DB 1; Length 634;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 281 gtttttaactttaaa 297
Db 462 GTTTTAACTTAAA 446

RESULT 4
US-08-450-595-1/C
; Sequence 1, Application US/08450595
; Patent No. 5798106
; GENERAL INFORMATION:
; APPLICANT: Schoenmakers, Johannes G
; APPLICANT: Koniings, Rudolph NH
; APPLICANT: Moelans, Inge IMD
; TITLE OF INVENTION: No. 5798106el protein
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Smithkline Beecham Corporate Patents -US
; STREET: UW2220, Po Box 1539
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/450,595
; FILING DATE: 25-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/949645
; FILING DATE: 04-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Jervis, Herbert H.
; REGISTRATION NUMBER: 31,171
; REFERENCE/DOCKET NUMBER: B2992
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-270-5065
; TELEFAX: 215-270-5090
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 634 base pairs
; TYPE: nucleic acid

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: PLASMODIUM
STRAIN: FALCIPARUM
IMMEDIATE SOURCE:
CLONE: 16K
US-08-450-595-1

Query Match 2.9%: Score 17; DB 1; Length 634;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 281 gtcttttaacttaaa 297
|||||
Db 462 GTTTTAACTTTAA 446

RESULT 5
US-08-007-775-3
Sequence 3, Application US/08007775
Patent No. 5340733
GENERAL INFORMATION:
APPLICANT: Takashi UENO et al.
TITLE OF INVENTION: MBOI RESTRICTION-MODIFICATION GENES
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wenderoth, Lind & Ponack
STREET: 805 Fifteenth Street, N.W., #700
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 500 kb
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/007,775
FILING DATE: 19930122
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Warren M. Cheek, Jr.
REGISTRATION NUMBER: 33,367
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-8650
TELEFAX:
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2659 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
HYPOTHETICAL:
ANTI-SENSE:
FRAGMENT TYPE:
ORIGINAL SOURCE:
ORGANISM: Moraxella bovis
STRAIN: 10900
INDIVIDUAL ISOLATE:
DEVELOPMENTAL STAGE:
HAPOTYPE:

TISSUE TYPE:
CELL TYPE:
CELL LINE:
ORGANELLE:
IMMEDIATE SOURCE:
LIBRARY:
CLONE:
POSITION IN GENOME:
CHROMOSOME/SEGMENT:
MAP POSITION:
UNITS:
FEATURE:
NAME/KEY:
LOCATION: 12
IDENTIFICATION METHOD:
OTHER INFORMATION: /note= "inosine"
PUBLICATION INFORMATION:
AUTHORS:
TITLE:
JOURNAL:
VOLUME:
ISSUE:
PAGES:
DATE:
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO:
US-08-007-775-3

Query Match 2.9%: Score 17; DB 1; Length 2659;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 294 taaacatcaaaaatt 310
|||||
Db 1208 TAAACATCAAAAATT 1224

RESULT 6
US-08-805-918-1/c
Sequence 1, Application US/08805918
Patent No. 5885821
GENERAL INFORMATION:
APPLICANT: MAGOTA, Koji
APPLICANT: MASUDA, Toyofumi
APPLICANT: SUZUKI, Yuji
APPLICANT: YABUTA, Masayuki
TITLE OF INVENTION: PROCESS FOR PRODUCTION OF SECRETORY KEY2
TITLE OF INVENTION: DERIVATIVES
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/805,918
FILING DATE: 04-MAR-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 8-073217
FILING DATE: 04-MAR-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 8-352580

FILING DATE: 16-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Meuth, Donna M.
REGISTRATION NUMBER: 36,607
REFERENCE/DOCKET NUMBER: 001560-295
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-6620
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2848 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Saccharomyces cerevisiae
STRAIN: X2180-1B
FEATURE:
NAME/KEY: CDS
LOCATION: 170..2611
US-08-805-918-1

Query Match 2.9%; Score 17; DB 2; Length 2848;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 331 tagtagttataattat 347
Db 157 TAGTAGTTATATATAT 141

RESULT 7
US-08-464-402-1
Sequence 1, Application US/08464402
Patent No. 5858705
GENERAL INFORMATION:
APPLICANT: WEI, ET AL.
TITLE OF INVENTION: Human DNA Ligase III
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
ADDRESSEE: CECCHI, STEWART & OLSTEIN
STREET: 6, BECKER FARM ROAD
CITY: ROSELAND
STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/464,402
FILING DATE: June 5, 1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/03939
FILING DATE: 31 MAR 95
ATTORNEY/AGENT INFORMATION:
NAME: FERRARO, GREGORY D.
REGISTRATION NUMBER: 36,134
REFERENCE/DOCKET NUMBER: 325800-388
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3417 BASE PAIRS
TYPE: NUCLEIC ACID
STRANDEDNESS: SINGLE

TOPOLOGY: LINEAR
MOLECULE TYPE: cDNA
US-08-464-402-1

Query Match 2.9%; Score 17; DB 2; Length 3417;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 232 tgcagttaaaaaaaaa 248
Db 3394 TGCAGTTAAAAAAA 3410

RESULT 8
US-09-054-775C-1
Sequence 1, Application US/09054775C
Patent No. 6284504
GENERAL INFORMATION:
APPLICANT: Wei, Yang-Fei
Yu, Guo-Liang
Haseltine, William
TITLE OF INVENTION: Human DNA Ligase III
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: MD
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/054,775C
FILING DATE: 03-Apr-1998
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/464,402
FILING DATE: 05-JUN-1995
APPLICATION NUMBER: PCT/US95/03939
FILING DATE: 31-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Hoover, Kenley K.
REGISTRATION NUMBER: 40,302
REFERENCE/DOCKET NUMBER: PFI61D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-309-8504
TELEFAX: 301-309-8439
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3417 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-054-775C-1

Query Match 2.9%; Score 17; DB 4; Length 3417;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 232 tgcagttaaaaaaaaa 248
Db 3394 TGCAGTTAAAAAAA 3410

RESULT 9

US-08-702-572-14/C
; Sequence 14, Application US/08702572
; Patent No. 5965386
; GENERAL INFORMATION:
; APPLICANT: Kerry-Williams, Sean M
; APPLICANT: Gilbert, Sarah C
; TITLE OF INVENTION: Yeast Strains and Modified Albumins
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Centeon L.L.C.
; STREET: 1020 First Avenue
; CITY: King of Prussia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19406-1310
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Microsoft Word 6.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/702,572
; FILING DATE: 11-NOV-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: NO 95/23857
; FILING DATE: 1-MAR-1995
; APPLICATION NUMBER: GB 9404270.2
; FILING DATE: 5-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Naomi Biswas
; REGISTRATION NUMBER: 38,384
; REFERENCE/DOCKET NUMBER: CE0114 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610/878/4294
; TELEFAX: 610/878/4221
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4106 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Saccharomyces cerevisiae
; US-08-702-572-14

Query Match 2.9%; Score 17; DB 2; Length 4106;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 331 tagagattataattat 347
|||||
Db 1316 TAGAGATTATAATTAT 1300

RESULT 10
5521093-4/C
; Patent No. 5521093
; APPLICANT: LEMONE, YVES;NGUYEN, MARTINE;ACHSTETTER, TILMAN
; TITLE OF INVENTION: YEAST VECTOR CODING FOR HETEROLOGOUS
; GENE FUSIONS LINKED VIA KEX2 CLEAVAGE SITE AND CODING FOR
; TRUNCATED KEX2 GENES
; NUMBER OF SEQUENCES: 8
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/393,025
; FILING DATE: 23-FEB-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 191,354
; FILING DATE: 07-FEB-1994

APPLICATION NUMBER: 26,121
; FILING DATE: 04-MAR-1993
; APPLICATION NUMBER: 500,885
; FILING DATE: 29-MAR-1990
; SEQ ID NO:4
; LENGTH: 4732
5521093-4

Query Match 2.9%; Score 17; DB 6; Length 4732;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 331 tagagattataattat 347
|||||
Db 1336 TAGAGATTATAATTAT 1320

RESULT 11
US-09-305-639-4
; Sequence 4, Application US/09305639
; Patent No. 6200778
; GENERAL INFORMATION:
; APPLICANT: Treco, Douglas A.
; APPLICANT: Seiden, Richard F.
; TITLE OF INVENTION: GENOMIC SEQUENCES FOR PROTEIN PRODUCTION AND DELIVERY
; FILE REFERENCE: 07236/016001
; CURRENT APPLICATION NUMBER: US/09/305,639
; CURRENT FILING DATE: 1999-05-05
; EARLIER APPLICATION NUMBER: 60/084,663
; EARLIER FILING DATE: 1998-05-07
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 6038
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-305-639-4

Query Match 2.9%; Score 17; DB 4; Length 6038;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 435 gcagaatgtaaaagta 451
|||||
Db 1446 gcagaatgtaaaagta 1462

RESULT 12
US-09-305-639-1
; Sequence 1, Application US/09305639
; Patent No. 6200778
; GENERAL INFORMATION:
; APPLICANT: Treco, Douglas A.
; APPLICANT: Heattlein, Michael W.
; APPLICANT: Seiden, Richard F.
; TITLE OF INVENTION: GENOMIC SEQUENCES FOR PROTEIN PRODUCTION AND DELIVERY
; FILE REFERENCE: 07236/016001
; CURRENT APPLICATION NUMBER: US/09/305,639
; CURRENT FILING DATE: 1999-05-05
; EARLIER APPLICATION NUMBER: 60/084,663
; EARLIER FILING DATE: 1998-05-07
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 7622
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-305-639-1

Query Match 2.9%; Score 17; DB 4; Length 7622;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 435 gcagaatgtaaaagta 451
DB 1446 gcagaatgtaaaagta 1462

RESULT 13

US-08-835-728D-96/C
Sequence 96, Application US/08835728D
Patent No. 6017704
GENERAL INFORMATION:
APPLICANT: Herman, James G.
APPLICANT: Baylin, Stephen B.
TITLE OF INVENTION: Methylation Specific Detection
NUMBER OF SEQUENCES: 216
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 4225 Executive Square, Suite 1400
CITY: La Jolla
STATE: CA
COUNTRY: USA
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/835,728D
FILING DATE: April 11, 1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/656,716
FILING DATE: June 03, 1996,
ATTORNEY/AGENT INFORMATION:
NAME: Halle, Lisa A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 07265/125001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619/678-5070
TELEFAX: 619/678-5099
INFORMATION FOR SEQ ID NO: 96:
SEQUENCE CHARACTERISTICS:
LENGTH: 29 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-08-835-728D-96

Query Match 2.7%; Score 16; DB 3; Length 29;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 239 aaaaaaaaaaacact 254
DB 28 AAAAAAAAAAACACT 13

RESULT 14

US-08-835-728D-200
Sequence 200, Application US/08835728D
Patent No. 6017704
GENERAL INFORMATION:
APPLICANT: Herman, James G.
APPLICANT: Baylin, Stephen B.
TITLE OF INVENTION: Methylation Specific Detection
NUMBER OF SEQUENCES: 216
CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Richardson P.C.
STREET: 4225 Executive Square, Suite 1400
CITY: La Jolla
STATE: CA
COUNTRY: USA
ZIP: 92037

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/835,728D
FILING DATE: April 11, 1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/656,716
FILING DATE: June 03, 1996,
ATTORNEY/AGENT INFORMATION:
NAME: Halle, Lisa A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 07265/125001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619/678-5070
TELEFAX: 619/678-5099
INFORMATION FOR SEQ ID NO: 200:
SEQUENCE CHARACTERISTICS:
LENGTH: 29 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-08-835-728D-200

Query Match 2.7%; Score 16; DB 3; Length 29;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 239 aaaaaaaaaaacact 254
DB 2 AAAAAAAAAAACACT 17

RESULT 15
US-09-490-558-96/C
Sequence 96, Application US/09490558
Patent No. 6265171

GENERAL INFORMATION:
APPLICANT: Herman, James G.
APPLICANT: Baylin, Stephen B.
TITLE OF INVENTION: Methylation Specific Detection
NUMBER OF SEQUENCES: 216
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 4225 Executive Square, Suite 1400
CITY: La Jolla
STATE: CA
COUNTRY: USA
ZIP: 92037

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/490,558
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/835,728
FILING DATE:
ATTORNEY/AGENT INFORMATION:

NAME: Haile, Lisa A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 07265/125001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619/678-5070
TELEFAX: 619/678-5099
INFORMATION FOR SEQ ID NO: 96:
SEQUENCE CHARACTERISTICS:
LENGTH: 29 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-09-490-558-96

Query Match 2.7%; Score 16; DB 4; Length 29;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 239 aaaaaaaaaaacct 254
|||||
Db 28 AAAAAAAAAACAACT 13

Search completed: May 22, 2002, 08:31:22
Job time: 7007 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 22, 2002, 07:30:35 ; Search time 2968.03 Seconds
(without alignments)
2696.635 Million cell updates/sec

Title: US-09-693-205-7_COPY_12200_12792

Sequence: 1 acatcttctgttaccagctc.....acctccacatttattgctt 593

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 13736207 seqs, 674847542 residues

Word size : 0

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

EST:*
1: em_estda:*
2: em_esthum:*
3: em_estlin:*
4: em_estlinu:*
5: em_estlov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inu:*
15: em_gss_pln:*
16: em_gss_vitl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	423	71.3	500	9 AI803488	AI803488 tcl17902.x
C 2	419	70.7	543	9 AI932370	AI932370 wd27e11.x
C 3	413	69.6	828	10 BM470780	BM470780 AGENCOURT
C 4	382	64.4	660	10 BE890125	BE890125 601513104
C 5	373	62.9	497	9 AM087745	AM087745 xb68f08.x
C 6	368	62.1	368	9 AA683013	AA683013 aeb1b08.s
C 7	355	59.9	356	10 BE856736	BE856736 7f68a06.x
C 8	355	59.9	356	10 BF438152	BF438152 7g67f12.x
C 9	345	58.2	469	9 AI217518	AI217518 qh20q08.x
C 10	341	57.5	422	9 AI076834	AI076834 o246d05.x
C 11	330	55.6	410	9 AA954825	AA954825 oo39d10.s
C 12	310	52.3	528	9 AA417817	AA417817 zv04h08.t
C 13	290	48.9	416	9 AI49896	AI49896 tn97d12.x
C 14	273	46.0	404	9 AA809783	AA809783 nw66c04.s
C 15	240	40.5	456	9 AA417676	AA417676 zv04d08.r
C 16	233	39.3	632	9 AM968633	AM968633 ESR380709
C 17	218	36.8	407	9 AA481507	AA481507 aa34c04.s

C 18	208	35.1	739	9 AL567149	AL567149 AL567149
C 19	206	34.7	351	9 AM262498	AM262498 xg85c06.x
C 20	192.	32.4	441	10 NS9442	NS9442 yz30b12.s1
C 21	172	29.0	349	9 AI561086	AI561086 tq26b04.x
C 22	165	27.8	483	10 N48291	N48291 yy77c12.s1
C 23	164	27.7	287	10 T17045	T17045 NIB26 Norm
C 24	155	26.1	418	10 N46342	N46342 yy74a10.s1
C 25	130	21.9	1016	10 BE886315	BE886315 601439161
C 26	106	17.9	789	10 BE889418	BE889418 601512535
C 27	79	13.3	840	10 BF653898	BF653898 602082463
C 28	74	12.5	203	9 AA228047	AA228047 zt58a07.r
C 29	74	12.5	1079	10 BM476997	BM476997 AGENCOURT
C 30	70	11.8	727	10 BG619034	BG619034 602616564
C 31	46	7.8	236	9 AA228008	AA228008 zt58a07.s
C 32	41	6.9	712	9 AV716805	AV716805 AV716805
C 33	41	6.9	818	10 BI599024	BI599024 603250412
C 34	32	5.4	162	10 BE894617	BE894617 601433235
C 35	29	4.9	478	10 BI871763	BI871763 603393147
C 36	28	4.7	677	10 BF576832	BF576832 602132834
C 37	26	4.4	201	9 AV167991	AV167991 AV167991
C 38	26	4.4	230	9 BB297230	BB297230 BB297230
C 39	26	4.4	230	9 AI447763	AI447763 mg82f11.x
C 40	26	4.4	302	9 AI865982	AI865982 wk88903.x
C 41	26	4.4	428	10 BE862199	BE862199 UI-M-BH0-
C 42	26	4.4	460	9 AI506773	AI506773 vms8g10.x
C 43	26	4.4	530	10 BM18416	BM18416 LO910P10-
C 44	26	4.4	543	9 AM553584	AM553584 L0255807-
C 45	26	4.4	566	10 BM119091	BM119091 L0920E06-

ALIGNMENTS

RESULT 1	AI803488/c				
LOCUS	AI803488	500 bp	mRNA	linear	EST 13-DEC-1999
DEFINITION	tcl17902.x1 Soares_NHMPU_S1 Homo sapiens cDNA clone IMAGE:2064146				
	3', mRNA sequence.				
ACCESSION	AI803488				
VERSION	AI803488.1	GI:5368882			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.				
REFERENCE	1 (bases 1 to 500)				
AUTHORS	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .				
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index				
JOURNAL	Unpublished (1997)				
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgaps-remail.nih.gov This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Insert Length: 1089 Std Error: 0.00 Seq primer: -400P from GIBCO High quality sequence stop: 447. Location/Qualifiers 1..500				
FEATURES					
source					

1..500
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2064146"
/tissue_type="Soares_NHMPU_S1"
/tissue_type="Pooled human melanocyte, fetal heart, and pregnant uterus"
/lab_host="DH10B"
/note="Organ: mixed (see below); Vector: pRT730-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (melanocyte 2NbM, pregnant uterus NbMpu, and fetal heart NBH19M) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization

DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov

Plate: LLAM1218 row: k column: 16
High quality sequence stop: 680.

FEATURES

source

1. 828
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:533575"
/clone_1lb="NIH_MGC_71"
/issue_type="leiomysarcoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: uterus; Vector: pCMV-Sport6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 2.1 kb.
BASE COUNT 262 a 107 c 138 g 321 t
ORIGIN

Query Match

Best Local Similarity 69.6%; Score 413; DB 10; Length 828;
Matches 593; Conservative 0; Mismatches 0; Indels 2; Gaps 1;

QY 1 acatctatggttacaaggtcctctgtttgatgaagatagcaaggaactcaaatggt 60
|||||
DB 143 ACATCTTATGTTTACAGGCTTCTGTTGATGAGATGACAGGAAACTCAAAATGCT 202
QY 61 ggcagttctatcacagttgttagtattgtcttgcgaactgcttcgaagaacatt 120
|||||
DB 203 GGCAGTCTTATACCAAGTTGATATGTTCTTGGAACCTGCTGCCAAGCAACATT 262
QY 121 tattaactgttagaacactgtcttattgttggtagacattttccacaatggtata 180
|||||
DB 263 TATTACTGTTAGAACACTGCTTATGTTGATGATATTTCCACAAATGTTATA 322
QY 181 attatatagttggttgaacagatgcaactcttctgttctcaaggtgctcagtt-- 238
|||||
DB 323 ATTTATATAGTGTGTGAACAGATGCAATCTTTGTTCTTAAAGTCTGCAGTTAA 382
QY 239 aaaaaaaacaaccttcttccaatgcatgtcagtgagtttttcaacttaaa 298
|||||
DB 383 AAAAAAACAACCTTTCTTCAATATGCGCATGTAGTGAATTTTAACTTTAAA 442
QY 299 acatcaaaaatgtttaaatacatgttctcagtagttataatcatcgactatatt 358
|||||
DB 443 ACATCAAAAAATGTTAAATCATTTGTTATCTAGTAGTTTATTAATATCGGCTTATATT 502
QY 359 tcccctatgatagtcgaactgacatttaattcaatgtttctcgcgacatgcttact 418
|||||
DB 503 TCCCCATGATGATGCAAGATGACATTTAATGATGTTTCTCGCATCTCTTACT 562
QY 419 ttaacatattctcttcgcaaatgttaaaggtaataatagttatataagttact 478
|||||
DB 563 TTATACATATTTCTTTGCAAAATGTAAGGTATATATTTAGTTTATTAAGTACT 622
QY 479 ggcgtgaatgatactaaatatacttaataatgcaatgaaggtctacagaacatgttgaac 538
|||||
DB 623 GGCTGTAATGATGTAATATATCTTTATGCAATTAAGGCTTACAGAACATGTTGAAC 682
QY 539 ttttttactttatctgggaataaggaatgttgcacctccacatttattgctt 593
|||||
DB 683 TTTTCTTACTTTTATTTGGAATGAAGATGTTGACCTCCACATTTTATTTGCTT 737

RESULT 4

BE890125 660 bp mRNA linear EST 20-OCT-2000
LOCUS BE890125
DEFINITION 6013104F1 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:3914521 5',
mRNA sequence.
ACCESSION BE890125
VERSION BE890125.1 GI:10348134

KEYWORDS

EST.
human.

SOURCE

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 660)

NIH-MGC http://imgc.ncl.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: c9abds-remail.nih.gov

Tissue Procurement: ATCC

CDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLAM9736 row: g column: 02
High quality sequence stop: 628.

FEATURES

source

1. 660
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3914521"
/clone_1lb="NIH_MGC_71"
/issue_type="leiomysarcoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: uterus; Vector: pCMV-Sport6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 2.1 kb.
BASE COUNT 196 a 95 c 107 g 262 t
ORIGIN

Query Match

Best Local Similarity 64.4%; Score 382; DB 10; Length 660;
Matches 552; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 acatctatggttacaaggtcctctgtttgatgaagatagcaaggaactcaaatggt 60
|||||
DB 41 ACATCTTATGTTTACAGGCTTCTGTTGATGAGATGACAGGAAACTCAAAATGCT 100
QY 61 ggcagttctatcacagttgttagtattgtcttgcgaactgcttcgaagaacatt 120
|||||
DB 101 GGCAGTCTTATACCAAGTTGATATGTTCTTGGAACCTGCTGCCAAGCAACATT 160
QY 121 tattaactgttagaacactgtcttattgttggtagacattttccacaatggtata 180
|||||
DB 161 TATTACTGTTAGAACACTGCTTATGTTGATGATATTTCCACAAATGTTATA 220
QY 181 attatatagttggttgaacagatgcaactcttctgttctcaaggtgctcagtt--a 239
|||||
DB 221 ATTTATATAGTGTGTGAACAGATGCAATCTTTGTTGATGATGTTTAAAGTCTGCAAGTTAA 280
QY 240 aaaaaaaacaaccttcttccaatgcatgtcagtgagtttttcaacttaaa 299
|||||
DB 281 AAAAAAACAACCTTTCTTCAATATGCGCATGTAGTGAATTTTAACTTTAAA 340
QY 300 catcaaaaatggttaaatacatgttctcagtagttataatgcaatgcttactt 359
|||||
DB 341 CATCAAAAATGTTAAATATCTTTGTTATCTAGTAGTTTATTAATGCTTATATT 400
QY 360 ccccatgatagtcgaactgacatttaattcaatgttctcgcgacatgcttactt 419
|||||
DB 401 CCCCATGAATGATGCAAGTACATTTAATGATGTTTCTCGCATGCTCTTACTT 460
QY 420 taacatattctcttcgcaaatgttaaaggtaataatagttatataagttactg 479
|||||
DB 461 TAAATATATTTCTTTGCAAAATGTAAGGTATATATTTAGTTTATTAAGTACTG 520
QY 480 gctgtaaatgatacttaacttcttgaatgaaggtctacagaacatgttgaact 539
|||||

Db 521 GGTGTAATGATGCTAAATTAATTATGCAATTAAAGGGCTTACAGACATGTTGAACCT 580
 QY 540 tctttactttat 553
 Db 581 TTTTCTACTTTTAT 594

RESULT 5
 AM087745/c 497 bp mRNA linear EST 15-OCT-1999
 LOCUS IMAGE:2581479.3', mRNA sequence.
 DEFINITION X868f08.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
 ACCESSION AM087745
 VERSION AM087745.1 GI:6043550
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1 (bases 1 to 497)
 NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
 AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 TITLE Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Seq primer: -40UP from Gibco
 High quality sequence stop: 455.
 Location/Qualifiers
 1..497
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2581479"
 /clone_lib="Soares_NFL_T_GBC_S1"
 /lab_host="DH10B"
 /note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with
 a modified polylinker; Site: 1: Not I; Site 2: Eco RI;
 Equal amounts of plasmid DNA from three normalized
 libraries (fetal lung Ndh119W, testis NHT, and B-cell
 NCI-CGAP-GCB1) were mixed, and ss circles were made in
 vitro. Following HAP purification, this DNA was used as
 tracer in a subtractive hybridization reaction. The driver
 was PCR-amplified cDNAs from pools of 5,000 clones made
 from the same 3 libraries. The pools consisted of
 I.M.A.G.E. clones 297480-302087, 682632-687239,
 726408-728711, and 729096-731399. Subtraction by Bento
 Soares and M. Fatima Bonaldo."
 BASE COUNT 195 a 80 c 63 g 159 t
 ORIGIN

Query Match 62.1%; Score 373; DB 9; Length 497;
 Best Local Similarity 100.0%; Pred. No. 5.2e-134;
 Matches 373; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 172 aatgtatattatattatgtgtgtgaacagatcattttgttctaaagtgc 231
 Db 497 AAGTTATATATTAATATAGTGTGTTGAACAGATCAATCTTTGTCTAAAGGTGC 438
 QY 232 tgcagtaaaaaaaacacacaccccttcttcaatagatgagatgttttctaac 291
 Db 437 TGCAGTAAAAAACAACACCTTTCTTTCATATGCGATGAGGAGACTTTTAAAC 378
 QY 292 tttaaaacatcaaaaatgttaaaatcattgtgtatcctagtagttataatcgcg 351
 Db 377 TTTAAAAACATCAAAATTTGTTAAATCATTTGTTATCTAGTAGTTAATATACGCG 318
 QY 352 ttatattcccatgatgatcagaactgacattatcattgttgcgcgacgctt 411
 Db 317 TTTATATTTCCCATGATGATGACAGACTTATTAATTCATGTTGTCTCGCCATGCTT 258

QY 412 cttacttacaatattcttcttgacgaatgtaaaagtaagataatagattataaa 471
 Db 257 CTTTACTTTAACADATTTCTTTTGCAAGATGTAAGGTAATGATATGTTATATAA 198
 QY 472 gtgtactgctgtaaatgataatgataatcattatgcaatgaaggcttaagaatg 531
 Db 197 GTGTACTGGCTGTAATGATGATTAATTAATGCAATTAAAGGGCTTACAGAACATG 138
 QY 532 ttgaacttttt 544
 Db 137 TTGAACCTTTT 125

RESULT 6
 AA683013/c 368 bp mRNA linear EST 15-DEC-1997
 LOCUS IMAGE:970551.3', mRNA sequence.
 DEFINITION ae81b08.s1 Stratagene schizo brain S11 Homo sapiens cDNA clone
 ACCESSION AA683013
 VERSION AA683013.1 GI:2668904
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1 (bases 1 to 368)
 Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,
 Krizman, D., Kucab, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin,
 J., Moore, B., Schellenger, K., Steptoe, M., Tan, F., Theising, B.,
 White, Y., Wylie, T., Waterston, R., and Wilson, R.
 WashU-NCI human EST Project
 Unpublished (1997)
 TITLE JOURNAL
 COMMENT Contact: Wilison RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Possible reversed clone: polyT not found
 Seq primer: -40m13 fwd. ET from Amersham
 High quality sequence stop: 358.
 Location/Qualifiers
 1..368
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:970551"
 /clone_lib="Stratagene schizo brain S11"
 /sex="male"
 /tissue="schizophrenic brain S-11 frontal lobe"
 /dev_stage="34 years old"
 /lab_host="SOLR (kanamycin resistant)"
 /note="Vector: Bluescript SK-; Site: 1: EcoRI; Library
 constructed from S-11 frontal lobe, male, 34 years old,
 50% caucasian, 50% Aleutian. Schizophrenic suicide.
 Random primed into EcoRI site of ZAP II Vector. Mass
 excised. Custom library. Avg insert length 1.4kb.
 Material obtained by Johnston N., Torrey, E.F., Yolken R.,
 and the Stanley Neuropathology Consortium - Analysis of
 RNAs from the Brains of individuals with psychiatric
 Diseases (Unpublished) Stanley Neuropathology Laboratory,
 Johns Hopkins School of Medicine, Baltimore MD."

BASE COUNT 140 a 61 c 52 g 115 t
 ORIGIN

Query Match 62.1%; Score 368; DB 9; Length 368;
 Best Local Similarity 100.0%; Pred. No. 5.3e-132;
 Matches 368; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 35 gatagaacggaanaactcaaaatggtgagcttctattaccagttgttagttcttc 94
 Db 35 gatagaacggaanaactcaaaatggtgagcttctattaccagttgttagttcttc 94

Db 368 GATAGCAGCGAAGAACTCAAAATGTCGACCTTCTATTACCACTGTGTAGATTGTTTC 309
Qy 95 tggaaactgcttgcgaagaacaattatcaactgcttagaacactgtcttatgttctg 154
Db 308 TGGAAATCGCTTGGCAAGACACATTTATTACTGTTAGAACACTTGGCTTTATGTTGTG 249
Qy 155 tgcatacatctccacaatgltlaattatcatagtggtgtgaacaggaatgcacatc 214
Db 248 TGTCATATTTTCCACAATGTTATTATTATAGTGTGTTGAACGATGCAATGTT 189
Qy 215 ttgtgtcctaaaggctgctgaagttaaaaaacaacaccttcttccaatagtcagt 274
Db 188 TTTGTTCTTAAAGGTGCTGCAAGTTAAAAAACAACCTTTCTTCAATATGCGACAGT 129
Qy 275 agtggaaatttttaacttaaacatcaaaattgtttaaatcatgtgttactagt 334
Db 128 AGTGAAGATTTTAACTTTAAACATCAAAAATGTTAAATCATGTTTATCTAGT 69
Qy 335 agttataatcatcgctatattcccatgaatgatacagaactgaacattatcatg 394
Db 68 AGTTATTAATTATGCGTTATATTCCCATGATGATGATCAACACTTATTAATCATG 9
Qy 395 ttgtgtctc 402
Db 8 TTTGCTCTC 1

RESULT 7
BE856736/c 536 bp mRNA linear EST 29-SEP-2000
LOCUS 7f68a06.x1 Soares_NSF_F8_9W_OT_PA_P_S1 Homo sapiens cDNA clone
DEFINITION IMAGE:329794 3', mRNA sequence.
ACCESSION BE856736
VERSION BE856736.1 GI:10370063
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 536)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Seq primer: -400P from Glibco
High quality sequence stop: 448.
Location/Qualifiers
1. 536
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:329794"
/clone_lib="Soares_NSF_F8_9W_OT_PA_P_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Equal amounts of plasmid DNA from five normalized
libraries were mixed, and ss circles were made in vitro.
Following HAP purification, this DNA was used as tracer in
a subtractive hybridization reaction. The driver was
PCR-amplified cDNAs from pools of 5,000 clones made from
the same 5 libraries. The pools consisted of the following
libraries and cloneids: Soares NB2HP pool 1:
309384-310919, 323208-325895 Soares NB2HP pool 1:
145032-147335, 147720-148103, 148872-149255, 15002 -
150407, 151176-152327 Soares NB2HP pool 1:
758280-760583, 772104-774407 Soares NBHPA pool 1:
304776-306311, 320136-322823, 326280-326653 Soares NBHOP
pool 1: 723720-726407, 739080-740999 Subtraction by Bento
Soares and M. Fatima Bonaldo."

BASE COUNT 209 a 85 c 68 g 174 t
ORIGIN
Query Match 59.9%; Score 355; DB 10; Length 536;
Best Local Similarity 100.0%; Pred. No. 4,5e-127;
Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 239 aaaaaaaaacaccttcttccaatagtcagtcagtgagtttttaacttaaa 298
Db 442 AAAAAAAAAACACCTTTCTTCAATATGGCATATAGGAGTTTCTTAACTTTAAA 383
Qy 299 acatcaaaatggttaaatcatcctgtatcatagtaagtttaataatcgctatalt 358
Db 382 ACATCAAAATTTGTTAAATGATTTGTATCTACTATTATTAATTATGCGCTATATT 323
Qy 359 tcccatgaatgatcagaactgacattaatcatgcttgcctgcacatgcttact 418
Db 322 TCCCATGAATGATCAGACAGTACATTTAATTCATGTTGTGCTCCCATGCTTACT 263
Qy 419 ttaacatattcttcttgcaagaatgaaaggtgaatgaatagttatagaagtact 478
Db 262 TTAACATATTTCTTTGCAAGATGTAAAGTATGATATATAGTTTATTAAGTACT 203
Qy 479 ggcgtlaatgatgctaataatacttctgaatgaagggcttaagaacatggtgaac 538
Db 202 GGCTTAATATGATGCTTAATATATTACTTTATGCAATTAAGGCTTACAGAACATGTTGAAC 143
Qy 539 ttttttactttattgggaatgaaggaatgttgcacctcacatttatgctt 593
Db 142 TTTTATTACTTTTATTTGGAATTAAGAAATGTTTGCACTCCACATTTATTTGCTT 88

RESULT 8
BF438152/c 536 bp mRNA linear EST 30-MAR-2001
LOCUS 7g67fi2.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:3703462 3',
DEFINITION mRNA sequence.
ACCESSION BF438152
VERSION BF438152.1 GI:11450669
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 536)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Christopher Moskalkuk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL, send email to:
info@image.lnl.gov
Seq primer: -400P from Glibco
High quality sequence stop: 481.
Location/Qualifiers
1. 536
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3703462"
/clone_lib="NCI_CGAP_Lu24"
/tissue_type="Carcinoid"
/lab_host="DH10B"
/note="Organ: Lung; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; Plasmid DNA from the normalized
library NCI_CGAP_Lu5 was prepared, and ss circles were

made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clones 1414920-1417991 and 1520904-1522439). Subtraction by Bento Soares and M. Fatima Bonaldo.

BASE COUNT 212 a 85 c 67 g 172 t
ORIGIN

Query Match 59.9%; Score 355; DB 10; Length 536;
Best Local Similarity 100.0%; Pred. No. 4.5e-127;
Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 239 aaaaaaaacacaccccttcttccatgatgcatgtgagatctttaaacttaaa 298
Db 439 AAAAAAAAAACACCTTTCTTCAATATGCGATGTAGTGGAGTTTAACTTAA 380
QY 299 acatcaaaatltgtaaatcatgtctatctagtagttacaataatcgactaat 358
Db 379 ACATCAAAATGTGTAATCATGTGTATAGTATTAATTAATGCGCTTATAT 320
QY 359 tcccataatgatcgaactgacattcaatcatgttctgcgcagatcttact 418
Db 319 TCCCATGATGATGCAAGCTGACATTTAATGATGTTCTCGCCATCTTACT 260
QY 419 ttaacatattcttgcagaatgtaaaagttaataatagttataatgact 478
Db 259 TTACATATTTCTTTTGCAAAATGTAAAGTAAATGATTAATTAATGAGTACT 200
QY 479 ggcgtgaatgacgtcaaatatcttcatgcaatgaaggcttacagaacatgtgaac 538
Db 199 GCGTGAATGATGCTAAATATATCTTATGCAATTAAGGCTTACAGAAATGTGAAC 140
QY 539 ttttttacttattatgggaataagaaagtgttcacccacatttattgctt 593
Db 139 TTTTACTTTTATGCGAATAGGATGTTTGCACCTCCACATTTTATGCTT 85

RESULT 9
LOCUS AI217518/c 469 bp mRNA linear EST I7-MAR-1999
DEFINITION qh20g08.x1 Soares_NFL_T_GBC_SI Homo sapiens CDNA clone
IMAGE:1845278 3', mRNA sequence.
ACCESSION AI217518
VERSION AI217518.1 GI:3797333
KEYWORDS EST.

SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 469)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncigap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
This clone is available royally-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 1085 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 456.

FEATURES
Location/Qualifiers
1..469
source

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1845278"
/clone_lib="Soares_NFL_T_GBC_SI"
/lab_host="DH10B"
/note="Organ: pooled; Vector: p7T73D-Pac (Pharmacia) with
a modified polylinker; Site:1: Not I; Site:2: Eco RI;
Equal amounts of plasmid DNA from three normalized

libraries (fetal lung NbHL19W, testis NHT, and B-cell NCI-CGAP CG81) were mixed, and ss circles were used in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 297480-302087, 682632-687239, 726408-728711, and 729096-731399. Subtraction by Bento Soares and M. Fatima Bonaldo.

BASE COUNT 184 a 71 c 63 g 151 t
ORIGIN

Query Match 58.2%; Score 345; DB 9; Length 469;
Best Local Similarity 100.0%; Pred. No. 3.5e-123;
Matches 345; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 249 caaccttcttccatgatgcatgtgagttctttaaacttaaaacatcaaaa 308
Db 428 CAACCTTTCTTCAATATGCGATGTAGTGGAGTTTAACTTAAACATCAAAA 369
QY 309 ttgttaaatcatgttctatctagtagttataatcgcgttatattcccatgaa 368
Db 368 TTGTAAATCATGTGTATGATGATTAATTAATGCGCTTAATTTCCCATGAA 309
QY 369 tgatcagaactgacattcaatcatgttctgcgcagatcttcaacttaacatatt 428
Db 308 TGATCAGAACTGACATTTAATTCATGTTGTCTGCCATGCTTTTAACTTAAT 249
QY 429 tctttgcagaatgtaaaagtgtaataatagttataatgactgctgtaaat 488
Db 248 TCTTTTCAGATGTAAAGTAAATGATTAATTAATTAATTAATGATGCTGTAAT 189
QY 489 gatgctaaatattatgcaatgaaggcttcacaaacatgttgaactttttact 548
Db 188 GATGCTAAATATATCTTATGCAATTAAGGCTTACAGAAATGTGAACCTTTTACT 129
QY 549 ttattgggaataagaaatgttgcacccacatttattgctt 593
Db 128 TTTATTTGGAATAGGATGTTTGCACCTCCACATTTTATGCTT 84

RESULT 10
LOCUS AI078834/c 422 bp mRNA linear EST 10-AUG-1998
DEFINITION q246g05.x1 Soares_NhHMPU_SI Homo sapiens CDNA clone IMAGE:1678377
IMAGE:167834
ACCESSION AI078834
VERSION AI078834.1 GI:3413141
KEYWORDS EST.

SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 422)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncigap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
This clone is available royally-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Seq primer: -40M13 fwd. ET from Amersham
High quality sequence stop: 361.

FEATURES
Location/Qualifiers
1..422
source

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1678377"
/clone_lib="Soares_NhHMPU_SI"
/tissue_type="Pooled human melanocyte, fetal heart, and
pregnant uterus"

/lab_host="DH10B"
 /note="Organ: mixed (see below); Vector: pT73D-Pac
 (pharmacia) with a modified polylinker site_1: Not I;
 site_2: Eco RI; Equal amounts of plasmid DNA from three
 normalized libraries (melanocyte 2NHM, pregnant uterus
 NBHPU, and fetal heart NBH19M) were mixed, and ss circles
 were made in vitro. Following HAP purification, this DNA
 was used as tracer in a subtractive hybridization
 reaction. The driver was PCR-amplified cDNAs from pools of
 5,000 clones made from the same 3 libraries. The pools
 consisted of I.M.A.G.E. clones 260232-265223,
 340488-345479, and 484488-489479."

BASE COUNT 169 a 64 c 54 g 135 t
 ORIGIN

Query Match 57.5%; Score 341; DB 9; Length 422;
 Best Local Similarity 100.0%; Pred. No. 1.3e-121;
 Matches 341; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 253 ctttttttcaatagcagatgtagtgaggttttttaactttaaaacatcaaatgtc 312
 Db 422 ctttttttcaatagcagatgtagtgaggttttttaactttaaaacatcaaatgtc 363
 QY 313 taaatcatgtgtatctagtagttagtataattacgagctataatcccatgaatgat 372
 Db 362 TAAATCATGTGTATCTAGTAGTATATATATCGGCTTATATTTCCCATGAAATGAT 303
 QY 373 cagaactgacatttaattcattgttctgcagatgcttcttaactttaacatattctt 432
 Db 302 CAGAACTGACATTTAATTCATGTTTGTCCGCAATGCTTTTACTTTAACAATTTCTT 243
 QY 433 ttgcagaatgtaaaaggtaattgataattgataatgaatgtactgctgtaaatg 492
 Db 242 TTTCAGAAATGTAAGGTAATGATTAATTAATTAATTAATTAATTAATTAATG 183
 QY 493 cttaatacttataatgaatgaagggttacaagacatgttgaaccttttttaacttta 552
 Db 182 CTTAAATTAATTTATGCAATTAAGGCTTACAGAACATGTTGAACCTTTTTCATCTT 123
 QY 553 ttgggaataaagaatgttgcacctccacatttattgtctt 593
 Db 122 TTGGGAATTAAGGAATGTTTGACACTCCACATTTTATGCTT 82

RESULT 11 410 bp mRNA linear EST 07-JUL-1998
 AA954825 0096d10.s1 NCI_CGAP_Kid6 Homo sapiens cDNA clone IMAGE:1574323 3',
 LOCUS mRNA sequence.
 DEFINITION
 AA954825
 VERSION AA954825.1 GI:3118520
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 410)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: c9abps-r@mail.nih.gov
 Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: Stratagene, Inc.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/ILNL at:
 www.bio.lnlnl.gov/db/seq/image/image.html
 Insert Length: 2008 Std Error: 0.00

Seq primer: -40ml3 fwd. ET from Amersham
 High quality sequence stop: 393.
 location/Qualifiers

FEATURES
 source 1..410
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:1574323"
 /clone_1lb="NCI_CGAP_Kid6"
 /sex="mixed"
 /tissue="kidney tumor"
 /lab_host="SOLR (kanamycin resistant)"
 /note="Organ: kidney; Vector: Bluescript SK-; Site_1:
 EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
 Oligo df. Pooled kidney tumors. 5' adaptor sequence: 5'
 GAATTCGGACGACG 3' 3' adaptor sequence: 5'
 CTCGAGTTTTTTTTTTTTTTT 3' Average insert size: 1.0 kb."

BASE COUNT 162 a 64 c 51 g 133 t
 ORIGIN

Query Match 55.6%; Score 330; DB 9; Length 410;
 Best Local Similarity 100.0%; Pred. No. 2.3e-117;
 Matches 330; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 264 atatgcatgtagtgaggttttttaactttaaaacatcaaatgtttaaatcatgt 323
 Db 410 ATATGGCATGAGTGAGGATTTTAACTTTAAACATCAAAATGTTTAAATCATGTG 351
 QY 324 tttatcatgtagtttaataattatcgcgctataattccccaatgaatgacatgaca 383
 Db 350 TGTATCATGAGTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 291
 QY 384 tttaattcatgttctgtcgcagatgcttcttaactttaacatattcttgcagaatg 443
 Db 290 TTTAATTCATGTTTGTCTCGCCAGCATGCTTCTTAAATTAATTTTTCGAGAAATG 231
 QY 444 aaaggttaatgataatgattatataatgttactgctgttaaatgataatgataatg 503
 Db 230 AAAAGTAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 171
 QY 504 ttatgcaatlaaaggcttacaagacatgttgaaccttttttaactttatgtggaataag 563
 Db 170 TTATGCAATTAAGGCTTACAGAACATGTTGAACCTTTTACTTTATTTGGCAATG 111
 QY 564 gaatgttgcacctccacatttattgtctt 593
 Db 110 GAATGTTTGACACTCCACATTTTATGCTT 81

RESULT 12 528 bp mRNA linear EST 02-MAR-1998
 AA417817 zV04h08.r1 Soares_NHMPU_S1 Homo sapiens cDNA clone IMAGE:752703 5',
 LOCUS similar to gb:J03464 PROCOLLAGEN ALPHA 2(1) CHAIN PRECURSOR (HUMAN)
 DEFINITION
 AA417817
 VERSION AA417817.1 GI:2079618
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 528)
 AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisler, G., Jost, S.,
 Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin
 J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theisling, B.,
 White, Y., Wyllie, T., Waterston, R. and Wilson, R.
 TITILE WashU-NCI human EST project
 JOURNAL Unpublished (1997)
 COMMENT Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800

Fax: 314 286 1810
Email: estewatson.wmrl.edu
This clone is available royalty-free through LNL; contact the
IMAE Consortium (info@mae.llnl.gov) for further information.
Insert Length: 312 Std Error: 0.00
Seq Primer: -28m3 rev2 ET from Amersham
High quality sequence stop: 455.

FEATURES

Source

```

/organism="Homo sapiens"
/db_xref="GDB:5975541"
/db_xref="taxon:9606"
/clone_image:752703"
/clone_id="Soares_NhNHPU_s1"
/tissue_type="Pooled human melanocyte, fetal heart, and
pregnant uterus"
/lab_host="DH10B"
/note="Organ: mixed (see below); Vector: p7T73D-Pac
(Pharmacia) with a modified polylinker; Site_1: Not I;
Site_2: Eco RI; Equal amounts of plasmid DNA from three
normalized libraries (melanocyte 2NbHM, pregnant uterus
NbHMU, and fetal heart NbHM19W) were mixed, and ss circles
were made in vitro. Following HAP purification, this DNA
was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from pools of
5,000 clones made from the same 3 libraries. The pools
consisted of I.M.A.G.E. clones 260232-265223,
340488-345479, and 484488-489479."

```

BASE COUNT	159 a	105 c	95 g	169 t
ORIGIN				

52.3%; Score 310; DB 9; Length 528;

OY	202	aggatgcaactcttggctgcctaaagggcgccagcttcaaaaaaacactctctt	261
Db	219	AGGATGCATCTTTTGTGTCTAAAGGCGTGCAGTTAAAAAAAACACCTTTCTTT	278
OY	262	caataatgcatgtagtggagttctttaaacttaaaaaacatcaaaaatgtttaaacaat	321
Db	279	CAAAATGGCAGTGTAGTGAGTTTTTTTAACTTTAAAAACATCAAAAAATGTAAAAATCAT	338
OY	322	tgtgttaactagtagttcaataatcagcgctatactcccatgaatgataagaacgca	381
Db	339	TGTGTATCTAGTAGTTATATATATATATGCGCTTATATTTCGCCATGAATGATCAGACTCA	398
OY	382	catttaatcgaatgtgtctgcgcacatgctctcttaacttaacaatacttcctttgcagaat	441
Db	399	CATTTAAATTCAMGGTTGCTCGCCAGTGGCTTTTACTTTAACATATTTCTTTTGCAGAT	458
OY	442	gtaaaaggctaatgataatagtttcaacagtgtaactgacgtgtaaatgctaaata	501
Db	459	GTAAGAAGTATGATAATATTACTTTATATATAGTGCTGCGTCTAATATGATGCTAAATATA	518
OY	502	cttaatacgcaa	511
Db	519	CTTATGCA	528

[illegible]

REFERENCE	AUTHORS	JOURNAL	COMMENT
1 (pages 1 to 416)	NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap . National Cancer Institute, Cancer Genome Anatomy Project (CGAP) Tumor Gene Index Unpublished (1997)		Contact: Robert Strausberg, Ph.D.

```

FEATURES
    source
        CDNA Library Preparation Life Technologies, Inc.
        DNA Sequencing by: Greg Lennon, Ph.D.
        Clone Sequencing by: Washington University Genome Sequencing Center
        found through the I.M.A.G.E. Consortium/LMLN at:
        www-bio.llnl.gov/bdrp/image/image.html
        Insert length: 1662      Std Error: 0.00
        Seq primer: ~40up from Gbooc
        High quality sequence stop: 395.
        Location/Qualifiers
            1..416
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:2177495"
            /clone_1ib="NCI-CGAP clone"
            /tissue_type="moderately-differentiated endometrial
            adenocarcinoma, 3 pooled tumors"
            /lab_host="DH10B"
            /note="Organ: uterus; Vector: pCMV-Sport6; Site: 1: SalI;
            Site: 2: NotI; Cloned unidirectionally. Primer: Oligo dt.
            Average insert size 1.85 kb. Life Technologies catalog #:
            11539-012"
BASE COUNT
161 a      64 c      52 g      139 t
ORIGIN

```

BASE COUNT	161 a	64 c	52 g	139 t
ORIGIN				

query Match	48.98;	Score 290;	DB 9;	Length 416;
Best Local Similarity	100.08;	Pred. No. 6.4e-102;		
Matches 290;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0

Oy	304	aaatctgtttaaacaatctgtgttctacagtagttataatcatatcggctatattcc	363
Db	374	AAAAATGTGTAAATCAATCTGTCTACTAGTAGTTATTAATATATGCGCTTAATTTCC	315
Oy	364	atgaatgalcagaactgacattaaatcatgctgttcctcgccatgctcttacttaac	423
Db	314	ATGATGTATCAGAACTGACATTTAATTCATGTCTTCTGCGACCTCTTACTTTAAC	255
Oy	424	atatctctttggcgaatgtaaaagtgatgtaataatgattatataatgtagtacctg	483
Db	254	ATATTCTTTTGGCAAAATGTAAGAAGTAAATGATTAATTAATTAAGTGTACGCGC	195
Oy	484	taaatgactcctaatacatacttcatlbgcaatlaagggtcttacagaacatgtlgaacctllt	543
Db	194	TAAATGATGCTAAATATATCTTATTCACATTTAAGGCGTTACAGAACATGTGAAACTTTT	135
Oy	544	ttactcttatctgggataagaatgctgttcgcacctccacattttatgtct	593
Db	134	TTACCTTTTAATTTGGGAATTAAGGATATTTTGGACCTCCACATTTTATTTGCTT	85

RESULT	14
AA809783/c	
LOCUS	AA809783
DEFINITION	nw66cc04.s1 NCI.GGAP_GCB1 Homo sapiens CDNA clone IMAGE:125158 3', mRNA linear EST 18-FEB-1998
ACCESSION	AA809783
VERSION	AA809783.1 GI:2879189
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE	1 (bases 1 to 404)

AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
JOURNAL Tumor Gene Index
COMMENT Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
Ph.D., Gerald Marti, M.D.
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldi, Ph.D.

CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/ldrp/image/image.html
Insert Length: 643 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 397.
Location/Qualifiers

FEATURES

source

1. 404
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1251558"
/tissue_type="germinal center B cell"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, IgD+),
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
(NCI) and Dr. Gerald Marti (CBER). CDNA synthesis was
primed with a Not I - Oligo(dT) primer
[5'-TGTACCAATCTGAGAGGGAGCGCCGCTCATTTTCTTTT-3',
1. Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldi."

BASE COUNT 160 a 62 c 51 g 130 t 1 others
ORIGIN

Query Match 46.0%; Score 273; DB 9; Length 404;
Best Local Similarity 99.7%; Pred. No. 2.4e-95;

Matches 323; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 270 catgtatgagtttttctaacttaaaacatcaaaatgttaaaatcatgtgtat 329

Db 404 CATGTACTGAGTTTCTTTTAACTTAAACATCAAAATGTTAAATCTTTGTAT 345

QY 330 ctgtagttataataatcgcgtatataatcccatgaatgatacagaactgacattat 389

Db 344 CTAGTACTTATATATATCGGCTTATATCCCATGATGATGAGAACTGACATTTAAT 285

QY 390 tcatgttgcctgcacatgcttcttctaataacatatctcttgcagaatgtaaaagg 449

Db 284 TCAATGTTGTCNCGCCATGCTTCTTAACTTAACTATTTCTTTGAGAAATGTAAGG 225

QY 450 taatgaataatgattatacaatgactgctgtaaatgataatgataataacttctgc 509

Db 224 TAATGATAATTAAGTTTATTAAGTACTGCTTAAATGATGCTAAATATACCTTTATGC 165

QY 510 aattaaggtctacagacatgtgaaacttttctaacttatttgggaataagaatgt 569

Db 164 AATTAAGGCTTACAGAACTGTTGAACCTTTTCTTACTTTTATTTGGAAATAGGAATGT 105

QY 570 ttgcacctcacactttatgtctt 593

Db 104 TTGCACCTCACATTTATTTGCTT 81

RESULT 15

AA417676 456 bp mRNA linear EST 02-MAR-1998
LOCUS 2V04408.r1 Soares_Nhhmpu.S1 Homo sapiens cDNA clone IMAGE:752655 5'
DEFINITION similar to gb:U03464 PROCOLLAGEN ALPHA 2(1) CHAIN PRECURSOR (HUMAN
)'; mRNA sequence.

ACCESSION AA417676
VERSION AA417676
KEYWORDS AA417676.1 GI:2079495
SOURCE EST.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE

AUTHORS

Hallier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,
Kritzman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Ma, M., Martin,
J., Moore, B., Schellenberg, R., Steptoe, M., Tan, F., Theisling, B.,
White, Y., Wyllie, T., Waterston, R. and Wilson, R.
Mashu-NCI human EST Project
Unpublished (1997)

JOURNAL

COMMENT

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu

This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (infoimage.llnl.gov) for further information.
Insert Length: 836 Std Error: 0.00
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 383.
Location/Qualifiers

FEATURES

source

1. 456
/organism="Homo sapiens"
/db_xref="GDB:5975490"
/db_xref="taxon:9606"
/clone="IMAGE:752655"
/tissue_type="Soares_Nhhmpu.S1"
/tissue_type="pooled human melanocyte, fetal heart, and
pregnant uterus"
/lab_host="DH10B"
/note="Organ: mixed (see below); Vector: pT73D-Pac
(Pharmacia) with a modified polylinker; Site_1: Not I;
Site_2: Eco RI; Equal amounts of plasmid DNA from three
normalized libraries (melanocyte 2Nbhm, pregnant uterus
Nbhuu, and fetal heart Nbhl19w) were mixed, and ss circles
were made in vitro. Following HAP purification, this DNA
was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from pools of
5,000 clones made from the same 3 libraries. The pools
consisted of I.M.A.G.E. clones 260232-265223,
340488-345479, and 484488-489479."

BASE COUNT 131 a 101 c 81 g 143 t
ORIGIN

Query Match 40.5%; Score 240; DB 9; Length 456;
Best Local Similarity 100.0%; Pred. No. 1.2e-92;

Matches 240; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 202 agagtgcaacttttctgtcttaaggtgctgcaatgtaaaacacacatttctt 261

Db 217 AGGATGCAATCTTTTGTGCTTAAAGGCTGCTGCACTTAAAAAACAACCTTTCTTT 276

QY 262 caataatgcatgtagtgaagtttttctaacttaaaacatcaaaatgttaaaatcat 321

Db 277 CAATATGCGCATGTAGTGAGTTTCTTTTAACTTAAACATCAAAATTTGTTAAATCAT 336

QY 322 tctgtatctagtattataataatcgcgtatataatcccatgaaatgatacagaatga 381

Db 337 TGTGTATCTAGTAGTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 396

QY 382 catttaattcgttctgcacatgcttcttctaacttaacatatcttcttgcagaat 441

Db 397 CATTTAATTCAGCTTGTCTCGCCATGCTTCTTACTTTACATATTTCTTTGCAGAA 456

Search completed: May 22, 2002, 07:30:39
Job time: 10809 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 22, 2002, 05:31:50 ; Search time 3328.52 Seconds

(without alignments)
3143.520 Million cell updates/sec

Title: US-09-693-205-7_COPY_1_500

Perfect score: 500
Sequence: 1 atgattacaggaagaccat.....gcagaattgtgtatcaaa 500

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1797656 seqs, 10463268293 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl:*
1: gb_da:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pt:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
------------	-------	-------------	--------	-------	-------------

1	500	100.0	12793	6	AX119931	Sequence
2	500	100.0	12793	9	AF193556	AF193556 Homo sapi
3	500	100.0	92693	9	AL157766	AL157766 Human DNA
4	499	99.8	99819	2	AC079761	AC079761 Homo sapi
5	41	8.2	174140	2	AC069017	AC069017 Mus muscu
6	35	7.0	11492	6	AX119933	AX119933 Sequence
7	35	7.0	11493	10	AF193557	AF193557 Mus muscu
8	22	4.4	39760	10	U29187	U29187 Mus musculus
9	21	4.2	10334	1	AE006329	AE006329 Lactococc
10	21	4.2	47573	3	AF030694	AF030694 Plasmodiu
11	21	4.2	130192	9	AL157832	AL157832 Human DNA
12	21	4.2	163721	2	AC099718	AC099718 Mus muscu
13	21	4.2	169234	9	AC018633	AC018633 Homo sapi
14	20	4.0	1295	3	AF034998	AF034998 Hyphantri
15	20	4.0	57205	2	AC107995	AC107995 Homo sapi
16	20	4.0	68654	2	AC105967	AC105967 Mus muscu
17	20	4.0	77836	2	AC022573	AC022573 Homo sapi
18	20	4.0	111714	2	AL359432	AL359432 Homo sapi
19	20	4.0	134977	2	AC002347	AC002347 Homo sapi
20	20	4.0	143914	2	AC022711	AC022711 Homo sapi
21	20	4.0	151982	2	AC009939	AC009939 Homo sapi
22	20	4.0	152529	2	AC008863	AC008863 Homo sapi
23	20	4.0	153665	2	AL359743	AL359743 Homo sapi
24	20	4.0	157864	2	AC034303	AC034303 Homo sapi
25	20	4.0	159865	2	AC011568	AC011568 Homo sapi
26	20	4.0	163989	2	AP002416	AP002416 Homo sapi
27	20	4.0	167619	9	AC012148	AC012148 Homo sapi
28	20	4.0	171076	9	AC090231	AC090231 Homo sapi
29	20	4.0	177510	2	AC091871	AC091871 Homo sapi
30	20	4.0	177951	2	AC074383	AC074383 Homo sapi
31	20	4.0	17969	9	AC068107	AC068107 Homo sapi
32	20	4.0	178255	9	AC019197	AC019197 Homo sapi
33	20	4.0	180915	9	HS495L19	AL442637 Homo sapi
34	20	4.0	187449	9	AC016450	AC016450 Homo sapi
35	20	4.0	188643	9	AC092422	AC092422 Homo sapi
36	20	4.0	189006	2	AC103467	AC103467 Rattus no
37	20	4.0	189157	2	AC011129	AC011129 Homo sapi
38	20	4.0	203793	2	AC016263	AC016263 Homo sapi
39	20	4.0	213296	2	AC067784	AC067784 Homo sapi
40	20	4.0	217250	2	AC018595	AC018595 Homo sapi
41	19	3.8	721	6	A43543	A43543 Sequence 11
42	19	3.8	724	11	G57985	G57985 SHGC-103784
43	19	3.8	746	8	LETRAS14	X51904 Lycopersico
44	19	3.8	1075	8	AF068875	AF068875 Glycine m
45	19	3.8	1076	8	AF068874	AF068874 Glycine c

ALIGNMENTS

RESULT 1
AX119931
LOCUS AX119931 12793 bp DNA linear PAR 11-MAY-2001
DEFINITION Sequence 1 from Patent WO0129266.
ACCESSION AX119931
VERSION AX119931.1 GI:14036678
KEYWORDS
SOURCE
ORGANISM Homo sapiens
human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
1 (bases 1 to 12793)
Hudson,T.J., Engert,J. and Richter,A.
Identification of arascs mutations and methods of use therefor
Patent: WO 0129266-A 1 26-APR-2001;
JOURNAL MCGILL UNIVERSITY (CA) : Hopital Sainte-Justine (CA)
location/Qualifiers

FEATURES
source
1..12793
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT	4163	a	2256	c	2487	g	3887	t
------------	------	---	------	---	------	---	------	---

Query Match 100.0%; Score 500; DB 6; Length 12793;
 Best Local Similarity 100.0%; Pred. No. 1,5e-256;
 Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

0y 1 atgattacagaagaacatctactcaagctcagcttctaataccagaagattgacg 60
 Db 1 ATGATTACAGAAGAACCATTACTACAGCTGACGCTTTAAATCCAGAACGATTTCACG 60
 0y 61 tctattcaagaagaatgaatacatctcgcctggcagagaagaattgattgtcattgta 120
 Db 61 TCTTATCAAGGAAGTAATCAATCATCTTCGCCCTGGCAGAGATGATTGTTCAAGTGA 120
 0y 121 tccatttgaagaacagaagaatccaccatctgttcaatgcttaagaagttggaataa 180
 Db 121 TCCATTTGATGAAAACAGAAATCCACCATTCTTTTCATGCTTAAAGATGTTGGAAAA 180
 0y 181 tcttatacatcttccagaagattgacttatttgaatgagatgccattatccccaag 240
 Db 181 TCTTATATACATTTTTCAGAGATTGACTTTATTGATGATGCCACTTATCCCCAG 240
 0y 241 aactactcagaaggaaggtcagaacatgtgtgaactcatatagaactcagaatccact 300
 Db 241 AACTATACTAGAGGAAGGTCAAGACATGTGTGAACCTTAGACTGAGATTCCATCGTT 300
 0y 301 aactcatttgaagcagatgaatccgaagcacagcttccagaattttgaagacattgaca 360
 Db 301 ACTCATTTTAGACGATGATTCGAAACAGACTTCAGAAATTTTTCAGACATTTGACA 360
 0y 361 aaactctggaagggttcttccataaaatagatgcatcatatacaacatccgcttataa 420
 Db 361 AAACTTGGAGGGTTTGTCTTAAAAAATTAGATGCATCTATACCAACATCCCTTATTAA 420
 0y 421 aaatatctctctaccattccagaagtcgtcttctgacgataaaggagaagatgcatt 480
 Db 421 AAAATATATCTCACTTACCACTTACCAAGTCTGTCTTTTTCAGATATGAGAGATGCCATT 480
 0y 481 gcagaatctgttaatcaaa 500
 Db 481 GCAGAAATGTGTATCAAA 500

RESULT 2
 AF193556 12793 bp DNA linear Pri 07-FEB-2000
 LOCUS Homo sapiens sacsin (SACS) gene, complete cds.
 DEFINITION AF193556
 ACCESSION AF193556.1 GI:6907041
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 12793)
 AUTHORS Engert,J.C., Berube,P., Mercier,J., Dore,C., Lepage,P., Ge,B.,
 Bouchard,J.P., Mathieu,J., Melancon,S.B., Schalling,M.,
 Lander,E.S., Morgan,K., Hudson,T.J. and Richter,A.
 LARSACS, a spastic ataxia common in northeastern Quebec, is caused
 by mutations in a new gene encoding an 11.5-kb ORF
 TITLE
 JOURNAL Nat. Genet. 24 (2), 120-125 (2000)
 MEDLINE 20120709
 REFERENCE 2 (bases 1 to 12793)
 AUTHORS Engert,J.C., Berube,P., Dore,C., Lepage,P., Ge,B., Hudson,T.J. and
 Richter,A.
 TITLE Direct Submission
 JOURNAL Submitted (08-OCT-1999) Genome Centre, Montreal General Hospital,
 1650 Cedar Ave., Montreal, QC H3G 1A4, Canada
 FEATURES
 source 1..12793
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="13"
 /map="between D13S232 and D13S292"

mRNA
 gene
 CDS

1..12793
 /gene="SACS"
 /product="sacsin"
 1..12793
 /gene="SACS"
 77..11566
 /gene="SACS"
 /note="molecular chaperone"
 /codon_start=1
 /product="sacsin"
 /protein_id="AAF31262.1"
 /db_xref="GI:6907042"
 /translation="MNTFPGRELIVOMYPEDENRNHPSVWLKMWKNTLYTHFSEDL
 TLEDMPILPRILEEGOCVFLRLRPSVLIDDESAPQLPEFLADIVOLGCVFL
 KLDASIOHPLIKKRYISPLPSAVLIOMKPELDLONITSLTPTHDALPKFLASL
 TDSSEKERITIDELAFKRINSSQGISSTYKLGCVLHHTATLPADRLSTISVID
 SSDEATIRLANMLTEOLTKTSLCLVLDLEINAFYSHEVYQLMWLENLSKNE
 NNVLWMLPLKFIQISOMVSAELEDPPDIEVLKDLCENEGTYFPSPVTSPTDL
 HSLROIGLKNEASLEKEDVYAKITEALQVACDPOVLLAKATLLVLYKNHTLL
 OSSEKMTLKKIKVPACKEREPPNPGSLVMKGDLCNLCAPDMCDVGHALLIGSLP
 LVESLHVNLKRALGIETKPSLAVLKHKIVYDWYSKTFSEDEYVFOHILLETYGF
 MHDHNEGKDSRALKPRVYATGKKPCPLAOLVIRPIDDLQPLVHNPKMAFHO
 LFKVGSIEBELTSDHISWYIQITLKSDDLSEDSKQNLHMLNTIKMNSQNPAS
 PNTVPV IHHSKNPSKLLMKPIHECCYCDIKVDLNDLEDSEVPIILVHEDIIPMTAE
 WLKVPCLSTRILINENMGFEQSGREPLIVRLIKNTLEEYPSVSDIEFKELQADANA
 TPCSFILDMRRMADIRENLDPGMAACHOPALMSFNNSQFSQSDPVNITRLEGSLRG
 EVDYKFGELGFNSYVHTIDPIIMSREFIMEDININISIKIKDPSGIKINMSK
 OKRLRKPNPKPFIIDVPCOLPLVLEAPVSYNGTIPRLSPRQOEAKPSVSSCY
 NTADITSLVDESFSGHRLITTOYSKMYLKYLLIEETNPSLADDTVIKSSSSK
 ANTVPLSVLKEAALMKTCSSNNKLPSDEKSSCTLIQITVEEHHVFRRLADQSP
 LFRGDDPDAALFEEMAKSGOQSKRPDSDELQKTECTYLLCMTDGEALKLSBSG
 RRLGLVPCGAVVOLSEIODQKMYKPHIGEVCYPLRIKTLGVHINGCFAVSNR
 KEIMTKDKRNMNTFMHVLKAYLVQVSLYRDATSGELMDYTYVAVWPPDVLHD
 DESYICQGFVEDIANGKRELTKVFSDGYSVSMKNVRLDQSLIKRDPVGAARKIF
 LKYLRKTSKNICAVELPSSVYLGFEBAKQILLENFTSEQFTSEVFPPIQDIEA
 ELRDLPLMTFVLEKVEDFSGLVRLTPICSLEHPLVLPRLHPEGVAKLPIKD
 GRFPYSTQDYDLNPIILIKVQLGNAKQDILMDMLERAVSAELKSDHVAACLRS
 ILLSLIDELKLRIDPRANDFAKQOTIRFLPLTPAGSLDMKNSKFEPTMFAVD
 LYTASHDIVLCLOPLINDNSHSFRCGSSVLAWEFLGLKLPYDVLVYNLQKFAV
 SYDQGITTYOENTINACYKYLHEALMONETITSLIDKLPSPFSILVENAYVDSKVS
 PHLNFEADPYLQLPNRYKNRRELPEYVIGVQSTVEDPFLVLESIDQRECKOTE
 ENFQLCRRTISSEGLSLIREKQDFCEKNYKILPLDNLMLLPKSKLYNDQCPKIKV
 KDTVAYYCHADIPREVAVKLGAVPRKHALERYASNVCTTIGTEFGOEKALYSKRS
 ILNAVYSEKEMLEKLDONADDAKATEICVDPDRQHPVDIRIDDMWAPJOGFALCVN
 NOPFTEDVVRGIQNGKTEGKPEGNPYKQVIGFNSVYITIDCPSEIISGNDLCTFDP
 HARVAPGATSTISGMPFRDLDQDFRQSDVLDITGHEFKLNDCTMPFPRNEMA
 KYSELSVYSPASBRMQLDKRSDGAEILMLNEMKSTICEIDKSGCALVLYSVK
 GKITDGDRLKRQFHASVYDSYTKRKQKLDIPVOOTITTMDEDESGNLTTLICNRS
 GFSMEKYSKSVISAHKNODITLFPFGGVAACITINYYKKPHAFGFLPLSTETGLPFI
 LSVLQNTPIHAYKDTLKFELFEPPYNRDLQDPLCLVAVLYLQLKRRYFGSDPT
 PNIQDGSASHAVIITWINNSTKTRPFDNITLODELQKADYNITTRKVAENY
 RKHLHLEIGFNLVYNCDETANLYCHCLDADIIPVYVPADIRSLPMLNVDKVIYNC
 GKLPCRLOQTNLKLHSLKLDVYCRKDAEENEIEVEGIPLLITIDSVLYQTFDARKPK
 FLTTYHELIPSRKDLFMNTLYLKYSNILLNCAVAVFDISSADLLSVLPEYTKS
 CKRWKNDPESKMLNAMHFISESVYKQDEETPTDIDVDTLDKVALMLGTGFTVS
 NLSCHNAVPEGVDLLPLSLMLIAVFPNAGQDFYALMLAGCLOLNLKICSQSDAFVP
 LISCHTANSEPTSLIKLALHWVOTSTPFAEKLVDNDFALIMYRNCNLMHMSDQDI
 KILKSLPCKKSISGRVYSIGKRGCTCYVLTKTSPSAVEKMTQSSASFLKXIKHKL
 YEIVGCVADPDLVLYLHLPLKLEMLSDAKLEHLIYLNKRLSSABEISEKLEQFLP
 LESLILIHANSRLQAKHFYDRTVAVEVMLPEKFLIPNDFEKLLEDILKRNHVF
 MTSWEFLNLGKLYLSQOOLQFAKEISVANYENNMSKETQNTMVDLHNIHOER
 MQLSGNFKLELILFPLCPERAPAEFIEFHQYOEVNCTLPLEFNGAVNPKFOC
 DYLOILMTSCPILPEKATPLISKBOGSDYQBOLEOVLNMLNANNIDPLKVIYNC
 RNICNTITLDEEMKTRAKVLYASTIEPLSAEKREPRFOALRGAFVWEDGKMLLAPB
 VVINLEYSEDFEPVLYKLPLELGTIHLQKHLGTDITISTKYVLEVLSIKFNSGKQ
 LDPNMRKIVRVSGLFRSLQDSVAKRDELDNVNLDLAPYRSLQGRRLKSSILVEDDE
 APHYKSRIOGNIGVOMLVLDISQYLGKHGFTKLIIMFQPLRRLRSLSSILEQDLE
 ETPKYCGGALCSLOGRLOLLSSDSFTGLIRIKKHENDNAFLNEAKSARLAKALR
 EGTAKVCEPKIOTLTVKGFENIPHSRSTFALRFQNAVALVILYIHSDDSDINPL
 ALAMTLKSKATMDLISDTSLYMLGSCNDIYIRGKRLDSGYKSYDSSEPKLDPMPGT
 PLPAIBIHTLLMDPNMNVYPGSYVGYLVDAEGBDLYGSTQPIYTYAIIIVOEEREDAD
 NSSFLGKIYQIDIGYSEKVIYSSLDLYKRSREBSOSSRSDAPSTPTSPTEFLTEGLR

BASE COUNT 4163 a 2256 c 2487 g 3887 t
ORIGIN
SIPLEFSGRESNHSKTSKHOSKPR.LKTVNSIPLEKTVNSVVEQDAWKLEDSERKK.IIRRL
YLKHPKRPKNEDNBIANEVEFKHLONEINLEPKOAF.LPNDNRASRPFSTISAFOSD
KYSORFRTYNSMNOEATSKHSKROOKNEKCSGSGOYTSOEFVPPFKSGVNGVRE
RWLROAANFNSAANDLHKNAENWCEPKYSTLALTAIAVARGSDKQVKTALAA
ACIIEYSQOQLEGLINDVHTLEAVYCVDSLKTTRY.PDLLPFPQIPNDRFTSEVAMRYNECT
ACIIKIKRNEQOQV"

Query Match	100.0%	Score 500.	DB 9:	Length 12793;
Best Local Similarity	100.0%	Prod. NO.1.5e-256;		
Matches 500;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY 1	atgatttcagaagaccatgtactcagctcgaagcttctaataccagaagattgcacg	60		
Db 1	ATGATTTCAGAGAACCATGTACTCAGCTCGAGCTTTAAATCCAGAACGATTGGCACG	60		
QY 61	tcttatcagaagaaagaaatgaaatcacattctggcctgycagagaagaaatgtatgttcaatgta	120		
Db 61	TCTTATCAAGAGAAAGTAATGAAATACATCTCGCGCTGCGAGAAATGATTGTTCATGTGTA	120		
QY 121	tcoatttgaatgaaaaacgaataaccccatctgttccatagcgttaagaatggtttgaaaaa	180		
Db 121	TCCTATTGATGAAAAACAAAAATCACCCCATCTGTTCATAGCGCTTAAGATGATTGGAAAAA	180		
QY 181	tcttatatacatttccagaagattgaccttatatttgatagatgcaccattatcccgag	240		
Db 181	TCTTATATACATTTTTCAGAGATTGTACTTTATTTTATGATGAGATGCCACTTATCCCGAG	240		
QY 241	aactatactagaagaagltcagacatgtgtgaaactaataagactcagaatccatcgt	300		
Db 241	AACCTACTAGAGAGAGGTCAAGACATGTTGGAACTATTAGACTCAGAGATTCCATCGTT	300		
QY 301	agtcattttagacgatgatctgtgaagcagcttcacgaatttttaagacagacattgtca	360		
Db 301	AGTCATTTTAAACGATGAACTGTAAGCGACAGCTTCCAAATTTTTCAGACGATTTGTGACA	360		
QY 361	aaacttggagaggttgccttaaaaaatagatgcatactatatacaacatccgcttataa	420		
Db 361	AAAATTTGGAGGCGTTGTCTTTAAAAATTTAGATGCATCTATACAAACATCCGCTTATTAA	420		
QY 421	aaaatataatcatccatcaccatcacaagtcgltgttttcagataaataatggaagaatgcatt	480		
Db 421	AAAATATATTCATTCACCATTAACCAAGTGCGTTTTGCAGATAAATGGAAGATGCCATT	480		
QY 481	gcagaaattgtgtcaatcaaa 500			
Db 481	GCAGAAATTTGTATATCAAA 500			
RESULT 3				
LOCUS AL157766/c				
DEFINITION	Human DNA sequence from clone RP11-40020 on chromosome			
ACCESSION	AL157766			
VERSION	AL157766.9			
KEYWORDS	HTG.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominiade; Homo.			
JOURNAL	1 (bases 1 to 92693)			
COMMENT	Direct Submission Submitted (11-APR-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk On Apr 12, 2001 this sequence version replaced gi:12709968. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the			

FEATURES

Source

```

variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em., EMBL, Sw.,
SWISSPROT, Tr., TREMBL, Wp., WormPEP, information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/projects/C\_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 13, constructed by the Sanger Centre Chromosome 13
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr13
Rp11-40020 is from the library RPCR-11.1 constructed by the group
of Piefer de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBAC3.6
IMPORTANT: This sequence is not the entire insert of clone
Rp11-40020 It may be shorter because we sequence overlapping
sections only once, except for a 100 base overlap.
The true left end of clone Rp11-760m1 is at 92594 in this sequence.
The true right end of clone Rp11-72p19 is at 100 in this sequence.
location/Qualifiers
1..92693
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="13"
/map="q12.11-12.2"
/clone="Rp11-40020"
/clone_1fb="RPC1-11.1"
2390..2485
/note="MIR repeat: matches 81..192 of consensus"
2562..2673
/note="MIR repeat: matches 68..183 of consensus"
3896..4201
/note="AluY repeat: matches 3..308 of consensus"
5122..5397
/note="MER46C repeat: matches 1..286 of consensus"
18986..19294
/note="AluY repeat: matches 1..310 of consensus"
19644..19873
/note="MER46A repeat: matches 1..235 of consensus"
20613..20912
/note="AluSg1 repeat: matches 1..300 of consensus"
23342..23651
/note="AluSg1 repeat: matches 1..309 of consensus"
24769..24891
/note="L2 repeat: matches 2554..2682 of consensus"
25871..26011
/note="L2 repeat: matches 2356..2495 of consensus"
26033..26109
/note="L2 repeat: matches 2601..2688 of consensus"
26245..26344
/note="L2 repeat: matches 2154..2255 of consensus"
26938..27096
/note="MIR repeat: matches 3..175 of consensus"
27150..27653
/note="L2 repeat: matches 1063..1644 of consensus"
28522..28891
/note="THE1B repeat: matches 1..364 of consensus"
29447..29834
/note="1LME3A repeat: matches 5787..6164 of consensus"
36098..36415
/note="AluX repeat: matches 1..308 of consensus"
37202..37414
/note="MIR repeat: matches 22..262 of consensus"
37963..38254

```

repeat_region	/note="Alusg repeat: matches 9. .301 of consensus"
repeat_region	39703. .3908
repeat_region	/note="Alusg repeat: matches 1. .306 of consensus"
repeat_region	33790. .40093
repeat_region	/note="Alusx repeat: matches 1. .304 of consensus"
repeat_region	40126. .40416
repeat_region	/note="Alusg repeat: matches 1. .292 of consensus"
repeat_region	40444. .40733
misc_feature	/note="Alusg repeat: matches 1. .292 of consensus"
repeat_region	41322. .41405
repeat_region	/note="Single clone region. Assembly confirmed by restriction digest data."
repeat_region	41541. .41788
repeat_region	/note="Alusg repeat: matches 1. .248 of consensus"
repeat_region	44790. .45101
repeat_region	/note="Alusg repeat: matches 1. .313 of consensus"
repeat_region	45261. .45312
repeat_region	/note="13 copies 4 mer tgtt 88% conserved"
repeat_region	45899. .46206
repeat_region	/note="Alu repeat: matches 1. .307 of consensus"
repeat_region	46754. .47052
repeat_region	/note="Alu repeat: matches 1. .298 of consensus"
repeat_region	47067. .47365
repeat_region	/note="Alu repeat: matches 1. .299 of consensus"
repeat_region	47477. .47873
repeat_region	/note="L1M10 repeat: matches 5950. .6322 of consensus"
repeat_region	47889. .48229
repeat_region	/note="Alusx repeat: matches 1. .312 of consensus"
repeat_region	49168. .49212
repeat_region	/note="Alu repeat: matches 85. .126 of consensus"
repeat_region	49620. .49693
repeat_region	/note="L2 repeat: matches 1685. .1757 of consensus"
repeat_region	50704. .51032
repeat_region	/note="AlusC repeat: matches 1. .308 of consensus"
misc_feature	52204. .53009
repeat_region	/note="CpG island"
repeat_region	53978. .54137
repeat_region	/note="Alusx repeat: matches 1. .160 of consensus"
repeat_region	54179. .54511
repeat_region	/note="L1M6 repeat: matches 5822. .6172 of consensus"
repeat_region	55685. .55949
repeat_region	/note="Alusx repeat: matches 1. .305 of consensus"
repeat_region	57331. .57390
repeat_region	/note="30 copies 2 mer ga 75% conserved"
repeat_region	57357. .57392
repeat_region	/note="9 copies 4 mer gaga 91% conserved"
repeat_region	57753. .57930
repeat_region	/note="MIR repeat: matches 82. .262 of consensus"
repeat_region	58260. .58389
repeat_region	/note="MIR repeat: matches 2. .153 of consensus"
repeat_region	58564. .58611
repeat_region	/note="24 copies 2 mer ca 93% conserved"
repeat_region	59350. .59533
repeat_region	/note="Alusg repeat: matches 129. .313 of consensus"
repeat_region	59922. .60223
repeat_region	/note="Alu repeat: matches 85. .299 of consensus"
repeat_region	61036. .61144
repeat_region	/note="L2 repeat: matches 2581. .2656 of consensus"
repeat_region	62008. .62187
repeat_region	/note="TIGER1 repeat: matches 2238. .2418 of consensus"
repeat_region	62188. .62316
repeat_region	/note="Alu repeat: matches 1. .129 of consensus"
repeat_region	62330. .62363
repeat_region	/note="Alu repeat: matches 261. .294 of consensus"
repeat_region	62362. .62565
repeat_region	/note="TIGER1 repeat: matches 1586. .1787 of consensus"
repeat_region	62566. .62665
repeat_region	/note="Alusg repeat: matches 1. .302 of consensus"
repeat_region	62866. .64385
repeat_region	/note="TIGER1 repeat: matches 46. .1586 of consensus"
repeat_region	64386. .64694
repeat_region	/note="Alu repeat: matches 1. .306 of consensus"

repeat_region	64695..64713	/note="TIGER1 repeat: matches 29. .46 of consensus"
repeat_region	65088..65395	/note="L1PB2 repeat: matches 5405. .5733 of consensus"
repeat_region	65396..65569	/note="L1PB2 repeat: matches 136. .309 of consensus"
repeat_region	65571..65640	/note="L1PB2 repeat: matches 5728. .5791 of consensus"
repeat_region	65696..65717	/note="11 copies 2 mer ta 100% conserved"
repeat_region	65725..66096	/note="L1PB2 repeat: matches 5789. .6155 of consensus"
repeat_region	66371..66410	/note="10 copies 4 mer tgtg 82% conserved"
repeat_region	67586..67886	/note="AluYb repeat: matches 1. .299 of consensus"
repeat_region	69748..69930	/note="MIR repeat: matches 6. .248 of consensus"
repeat_region	70957..71267	/note="AluY repeat: matches 1. .311 of consensus"
repeat_region	71279..71413	/note="MER1B repeat: matches 548. .680 of consensus"
repeat_region	71411..71737	/note="MER1A repeat: matches 47. .485 of consensus"
repeat_region	71760..72075	/note="AluX repeat: matches 1. .295 of consensus"
repeat_region	72145..72256	/note="MER11-internal repeat: matches 42. .175 of consensus"
repeat_region	72454..72865	/note="MER11-internal repeat: matches 332. .739 of consensus"
repeat_region	72873..73249	/note="MER11-internal repeat: matches 883. .1261 of consensus"
Query Match	100.0%;	Score 500; DB 9; Length 92693;
Best Local Similarity	100.0%;	Pred. No. 1,4e-256;
Matches 500; Conservative	0; Mismatches	0; Indels 0; Gaps 0;
1	atgatttcagaagaaccatgttaccagctcagctcagcttcataatccagaagattgcagc	60
18909	ATGATTTCAGAGAGACCATGTACTGACCTGCAGCTTTAAATCCAGACATTTGGCAGC	18650
61	tctltacagaagaatgaatgatacatcttgcgcctgcgcagagaatgattgttcaatgta	120
18849	TCTTATCAAGAGAGTAATGAAATGAACTCTGGCGCTGCAGAGAAATGATTGTTCAATGTA	18790
121	tccatttgcagaaacagagaatccaccatctgttccatgcttgaatgattgttggaaaaa	180
18789	TCCATTTCATGAAAAACAAGAAATCACCCATCTGTTTCATGAGGCTTAAGATGTTGGAAAAA	18730
181	tcttatatacatcttltagaagatttgaacttatttataagaatgagcaactatccccag	240
18729	TCTTATATATACATTTTTCAGAGATTTCATCTTATTTATATAGAGCCACTTATCCAG	18670
241	aactactagaagaagtcagacacatgctgtggaactcatltagactcagaattccatcgtt	300
18669	AACATTCATAGAGAGAGTGCAGACATGTGTGGAATCATTTAGACACAGATTCCATGCTT	18610
301	agtcattttagacgatcattgaaagcaagcttcagaattctttagacagaattgttaca	360
18609	AGTCATTTTAGACGATGTGAATCTGAAGCACACCTCCCAAAATTTTACGACGACATGTGACA	18550
361	aaaacttggagaggttgcctttaaanaattgatagcatactataaaccctgcattataa	420
18549	AAAACCTTGAGGGTGTGCTTAAAAAATTTAGATGCATCTTATACAACTCCGCTTATTAA	18490
421	aaaatatcatcaccattaccagaagtcgcttlttgagataatggaagaatgaccatt	480
18489	AAAAATATTCATTCACCATTAACCAAGTGTGCTTTTGACAGTAAATGGAGAGATGCCATT	18430
481	gcagaagaattgttaatcaaa 500	

Db 18429 GCAGAAATGTGTATCAAA 18410

RESULT 4
AC079761 99819 bp .DNA .linear HTG 10-SEP-2000
LOCUS Homo sapiens chromosome UNK clone RP11-143G17, *** SEQUENCING IN
DEFINITION PROGRESS ***, 44 unordered pieces.
ACCESSION AC079761 GI:10047966
VERSION AC079761.1
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 99819)
AUTHORS Waterston, R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 99819)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (10-SEP-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

COMMENT ----- Genome Center -----
Center: Washington University Genome Sequencing Center
Web site: <http://genome.wustl.edu/gsc/index.shtml>
Project Information -----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 44 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1137: contig of 1137 bp in length
* 1138 1237: gap of unknown length
* 1238 2538: contig of 1301 bp in length
* 2539 2639: gap of unknown length
* 2639 3976: contig of 1338 bp in length
* 3977 4076: gap of unknown length
* 4077 5355: contig of 1279 bp in length
* 5356 5456: gap of unknown length
* 5456 6757: contig of 1302 bp in length
* 6758 6857: gap of unknown length
* 6858 8570: contig of 1713 bp in length
* 8571 8670: gap of unknown length
* 8671 9925: contig of 1255 bp in length
* 9926 10025: gap of unknown length
* 10026 11427 11526: contig of 1401 bp in length
* 11427 13266: gap of unknown length
* 13267 13366: contig of 1740 bp in length
* 13367 13366: gap of unknown length
* 13367 14794: contig of 1428 bp in length
* 14795 14894: gap of unknown length
* 14895 16054: contig of 1160 bp in length
* 16055 16154: gap of unknown length
* 16155 17395: contig of 1241 bp in length
* 17396 17495: gap of unknown length
* 17496 19287: contig of 1792 bp in length
* 19288 19388: gap of unknown length
* 19389 21294: contig of 1907 bp in length
* 21295 21394: gap of unknown length
* 21395 22944: contig of 1550 bp in length
* 22945 23044: gap of unknown length
* 23045 24421: contig of 1377 bp in length
* 24422 24521: gap of unknown length
* 24522 25870: contig of 1349 bp in length
* 25871 25970: gap of unknown length

25971 27230: contig of 1260 bp in length
* 27231 27330: gap of unknown length
* 27331 28778: contig of 1448 bp in length
* 28779 28878: gap of unknown length
* 28879 30893: contig of 2015 bp in length
* 30894 30993: gap of unknown length
* 30994 32461 32460: contig of 1467 bp in length
* 32461 32560: gap of unknown length
* 32561 33984: contig of 1424 bp in length
* 33985 34084: gap of unknown length
* 33985 35285: contig of 1201 bp in length
* 35286 35385: gap of unknown length
* 35386 37184: contig of 1799 bp in length
* 37185 37284: gap of unknown length
* 37285 39172: contig of 1888 bp in length
* 39173 39272: gap of unknown length
* 39273 40874: contig of 1602 bp in length
* 40875 40974: gap of unknown length
* 40975 42893: contig of 1919 bp in length
* 42894 42993: gap of unknown length
* 42994 44384: contig of 1391 bp in length
* 44385 44484: gap of unknown length
* 44485 45999: contig of 1515 bp in length
* 46000 46099: gap of unknown length
* 46100 48669: contig of 2570 bp in length
* 48670 48768: gap of unknown length
* 48770 50798: contig of 2029 bp in length
* 50799 50898: gap of unknown length
* 50899 52809: contig of 1911 bp in length
* 52810 52909: gap of unknown length
* 52910 55127: contig of 2218 bp in length
* 55128 55227: gap of unknown length
* 55228 58087: contig of 2860 bp in length
* 58088 58187: gap of unknown length
* 58188 61004: contig of 2817 bp in length
* 61005 61104: gap of unknown length
* 61105 64185: contig of 3081 bp in length
* 64186 64285: gap of unknown length
* 64286 67105: contig of 2820 bp in length
* 67106 67205: gap of unknown length
* 67206 70837: contig of 3632 bp in length
* 70838 70937: gap of unknown length
* 70938 75937: contig of 4900 bp in length
* 75938 80452: gap of unknown length
* 80453 80552: gap of unknown length
* 80553 84661: contig of 4109 bp in length
* 84662 84761: gap of unknown length
* 84762 90543: contig of 5781 bp in length
* 90543 90642: gap of unknown length
* 90643 94348: contig of 3706 bp in length
* 94349 94449 99819: contig of 5371 bp in length.

FEATURES
source
1..99819 Location/Qualifiers
1..1137 /organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="UNK"
/clone="RP11-143G17"
1..1137 /note="assembly_name:Contig20"
1238..2538 /note="assembly_name:Contig23"
2639..3976 /note="assembly_name:Contig30"
4077..5355 /note="assembly_name:Contig35"
5456..6757 /note="assembly_name:Contig39"
6858..8570 /note="assembly_name:Contig40"
8671..9925 /note="assembly_name:Contig41"

```

misc_feature      10026..11426
                    /note="assembly_name:Contig42"
misc_feature      11527..13266
                    /note="assembly_name:Contig44"
misc_feature      13367..14794
                    /note="assembly_name:Contig47"
misc_feature      14895..16054
                    /note="assembly_name:Contig48"
misc_feature      16155..17395
                    /note="assembly_name:Contig49"
misc_feature      17496..19287
                    /note="assembly_name:Contig50"
misc_feature      19388..21294
                    /note="assembly_name:Contig51"
misc_feature      21395..22944
                    /note="assembly_name:Contig52"
misc_feature      23045..24421
                    /note="assembly_name:Contig53"
misc_feature      24522..25870
                    /note="assembly_name:Contig54"
misc_feature      25971..27230
                    /note="assembly_name:Contig55"
misc_feature      27331..28778
                    /note="assembly_name:Contig57"
misc_feature      28879..30893
                    /note="assembly_name:Contig58"
misc_feature      30994..32460
                    /note="assembly_name:Contig59"
misc_feature      32561..33984
                    /note="assembly_name:Contig60"
misc_feature      34085..35285
                    /note="assembly_name:Contig61"
misc_feature      35386..37184
                    /note="assembly_name:Contig62"
misc_feature      37285..39172
                    /note="assembly_name:Contig63"
misc_feature      39273..40874
                    /note="assembly_name:Contig64"
misc_feature      40975..42893
                    /note="assembly_name:Contig65"
misc_feature      42994..44384
                    /note="assembly_name:Contig66"
misc_feature      44485..45999
                    /note="assembly_name:Contig67"
misc_feature      46100..48669
                    /note="assembly_name:Contig68"
misc_feature      48770..50798
                    /note="assembly_name:Contig69"
misc_feature      50899..52809
                    /note="assembly_name:Contig70"
misc_feature      52910..55127
                    /note="assembly_name:Contig71"
misc_feature      55228..58087
                    /note="assembly_name:Contig72"
misc_feature      58188..61004
                    /note="assembly_name:Contig73"
misc_feature      61105..64185
                    /note="assembly_name:Contig74"
misc_feature      64286..67105
                    /note="assembly_name:Contig74"

```

Query Match 99.8%; Score 499; DB 2; Length 99819;
 Best Local Similarity 100.0%; Pred. No. 4; 8e-256;
 Matches 499; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 atgattacaagaagacatgtaactcagctgcagctcttaaatccagaacgattgcaag 60
    |||||||
DB 48099 ATGATTACAGGAACCACTGCTCAGCGACGCTCTTAATCCAGACGATTGCAAG 48158
QY 61 tcttatacaagaagtaatactatcttctgctgcagagaattgatttcaatgta 120
    |||||||
DB 48159 TCTTATCAAGAGATATGATATCTCTGCGCAGAGAAFTGATTTCTCAATGCTGA 48218
QY 121 tccatttgatgaacagaatcacccatctgttcatggttaagaatggtttgaaaaa 180

```

```

DB 48219 TCCATTGATGAAGAAACAGAAATCACCCATCTGTTTCATGCGTTAGATGCTTGGAAAA 48278
QY 181 tcttatacatcttttcaagaagattacttatttgaagaatgacacttccccc 240
    |||||||
DB 48279 TCTTATATATCAATTTTTCAGAGATTTGACTTATTTGATGAATGCCACTTATCCCCAG 48338
QY 241 aactatactagaagaagtgtaagacatgtgtggaactcattagaactcagaattccact 300
    |||||||
DB 48339 AACTATATCTAGAGAGAGTGTAGACATGTGTGGAACATCATTAGACTCAGATTCATCGTT 48398
QY 301 agtcattttagacatgtaactcgaagacagcttccagaatttttagcagacatgta 360
    |||||||
DB 48399 AGTCATTTTAGACATGAATGTAATCTGAAGCAGCTTCCAGATTTTGAACAGACTTGTACA 48458
QY 361 aaactctggaggtgttctcttaaaaaataagatgcatctatacaacatcggctttaa 420
    |||||||
DB 48459 AAAACTTGGAGGGTGTCTTCTTAAAAAATGATGCACTTATACACATCCGCTTTTAA 48518
QY 421 aaatatatcatcacatcacatcacagtgctgttttgcagataatgagaagaatgcccatt 480
    |||||||
DB 48519 AAAATATATTCATTCACCATATACCAAGTCTGTGTTTGCAATATATGAGAGATGCCATT 48578
QY 481 gcagaatgtgttaatcaa 499
    |||||||
DB 48579 GCAGAAATGTGTATATCAA 48597

```

```

RESULT 5
AC069017 174140 bp DNA linear HTG 12-SEP-2001
LOCUS Mus musculus clone MGS3-342116, WORKING DRAFT SEQUENCE, 5 unordered
DEFINITION
AC069017
VERSION AC069017.20 GI:15559167
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 174140)
REFERENCE
  Metzker,M.L., Lewis,L.R., Hume,J., Edwards,C., Harris,C.,
  Dederich,D., Thomas,S., Okwuonu,G., Carlson,C., Garner,T.,
  Addison,S., Pace,A., Williams,G., Bonin,D., Brooks,A., Brown,J.,
  Buhay,C., Bunac,C., Burkett,C., Chacko,J., Chen,G., Chen,Z.,
  Cox,C., Davis,C., Delgado,O., Ding,Y., Dugan-Rocha,S.,
  Fernandez,C., Ferraguto,D., Fortum-Tansey,J., Gill,R.,
  Gorrell,J.H., Gunaratne,P., Haller,G., Hernandez,J., Hognes,M.,
  Hosak,H., Hou,X., Huber,J., Jackson,L., Jia,Y., Kelly,J., Kelly,S.,
  Kovar,C., Liu,J., Liu,W., Louisedge,H., Lozada,R.J., Martin,R.,
  Massey,E., McLeod,M.P., Mei,G., Moore,S., Morgan,M., Morris,S.,
  Neal,D., Nelson,A., Nguyen,R., Nguyen,N., Ogih,M., Parish,B.,
  Perez,L., Reiter,D., Say,J., Shen,H., Vasquez,L., Wallington,S.,
  Williamson,A., Wrensford,G., Zhou,X., Bouck,J., Hodgson,A.,
  Muzny,D.M., Rives,M., Scherer,S., Sodergren,E., Weinstock,G.,
  Worley,K. and Gibbs,R.
  Direct Submission
  Unpublished
  2 (bases 1 to 174140)
REFERENCE
  Worley,K.C.
  Direct Submission
  Submitted (17-MAY-2000) Human Genome Sequencing Center, Department
  of Molecular and Human Genetics, Baylor College of Medicine, One
  Baylor Plaza, Houston, TX 77030, USA
  On Sep 12, 2001 this sequence version replaced gi:14787161.
  ----- Genome Center
  Center: Baylor College of Medicine
  Center code: BCM
  Web site: http://www.hgsc.bcm.tmc.edu/
  Contact: hgsc-help@bcm.tmc.edu
  ----- Project Information
  Center project name: MMAP
  Center clone name: MGS3-342116

```

COMMENT

----- Summary Statistics

Sequencing vector: M13, L08821
Chemistry: Dye-terminator Big Dye: 63% of reads
Chemistry: Dye-terminator Big Dye: 63% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 175525 bases at least Q40
Consensus quality: 179254 bases at least Q30
Consensus quality: 181014 bases at least Q20
Estimated insert size: 178322; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-1p estimation
Quality coverage: 7.4x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 86925: contig of 86925 bp in length
* 86926 87025: gap of unknown length
* 87026 136728: contig of 49703 bp in length
* 136729 136828: gap of unknown length
* 136829 159942: contig of 23114 bp in length
* 159943 160042: gap of unknown length
* 160043 168814: contig of 8772 bp in length
* 168815 168914: gap of unknown length
* 168915 174140: contig of 5226 bp in length.

FEATURES

Location/Qualifiers

1..174140
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="MGS3-342116"

BASE COUNT 51417 a 32907 c 34252 g 55160 t 404 others
ORIGIN

Query Match 8.2%; Score 41; DB 2; Length 174140;
Best Local Similarity 100.0%; Pred. No. 8.8e-11;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 caggagaccatgactgactgacgtcttaataccagaa 49
|||||
Db 55267 CAGGAGACCATGTACTACGCTGCAGCTCTTAATCCAGAA 55307

RESULT 6
AX119933 11492 bp DNA linear PAT 11-MAY-2001
LOCUS
DEFINITION Sequence 3 from Patent WO0129266.
ACCESSION AX119933
VERSION AX119933.1 GI:14036679

KEYWORDS
SOURCE
ORGANISM

house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 11492)
Hudson,T.J., Engert,J., and Richter,A.
Identification of arcs mutations and methods of use therefor
Patent: WO 0129266-A 3 26-APR-2001;
MCGILL UNIVERSITY (CA) : Hopital Sainte-Justine (CA)

FEATURES

Location/Qualifiers

1..11492
/organism="Mus musculus"
/db_xref="taxon:10090"

BASE COUNT 3599 a 2280 c 2387 g 3226 t
ORIGIN

Query Match 7.0%; Score 35; DB 6; Length 11492;
Best Local Similarity 100.0%; Pred. No. 1.5e-07;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 203 gattgacttatattgattgagatgccactatccc 237
|||||
Db 127 GATTGACTTATTGATGAGATGCCACTTATCCC 161

RESULT 7

AF193557

LOCUS

AF193557

DEFINITION

AF193557

ACCESSION

AF193557.1

VERSION

AF193557.1

KEYWORDS

GI:6907043

SOURCE

house mouse.

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.1 (bases 1 to 11493)
Engert,J.C., Berube,P., Mercier,J., Dore,C., Lepage,P., Ge,B.,
Bouchard,J.P., Mathieu,J., Melancon,S.B., Schalling,M.,
Lander,E.S., Morgan,K., Hudson,T.J., and Richter,A.
ARSACS, a spastic ataxia common in northeastern Quebec, is caused
by mutations in a new gene encoding an 11.5-kb ORF
Nat. Genet. 24 (2), 120-125 (2000)
2 (bases 1 to 11493)
Engert,J.C., Berube,P., Dore,C., Lepage,P., Ge,B., Hudson,T.J., and
Richter,A.
Direct Submission
Submitted (08-OCT-1999) Genome Centre, Montreal General Hospital,
1650 Cedar Ave., Montreal, QC H3G 1A4, Canada

Location/Qualifiers

1..11493
/organism="Mus musculus"
/db_xref="taxon:10090"
/product="sacsin"
/note="molecular chaperone"
/codon_start=1
/product_id="AF31263.1"
/protein_id="AF31263.1"
/translation="MNTFMPGRELIVYVQWYFSEEDKRRPSSLMLKWKVKNLYIHSEDL
TLPDEMLIPRLTINEDQCVELIRLIPSVLLIDDETEAOLPEPLADYOKLGLYL
KRLDTSIQHPLVYKTIHSPILPSAIIQIMKIPLOKLCNQASLPLPHKDLRRFLASL
TDTSEKERRIQELTIFKRIHNSDDQISSYTKKGGKVLADHTAKLPTDRLSVYSD
SSDEATRIANMLKIEKLTSSCLKEVLKIDIGAFYQOEETOLMLLENLSLKE
NSNVLDIMPLKFIHMSOGHVAAGDFDDIDYRLDFNEEACPPYIFVSPDL
HSIRQIGLKRESSLAKEDYQVARKITAELOVSSCONDVAKKAKTLLVLNKNQSL
QSSGEMALKIKYVPAKCRPPNYRSGSLYKVDICLCAKPPMKDAHNVIVGSSILP
LVSEVHNLEQALSIFFKPTINVLAKFKYVDMYVYSKTSDEYVQFHVLIETIGF
MDHLSEGGKDFKALFPWYWGKNCPLQAVIKPHDLDLPYVIVNRYKTAHQ
LFRACSGISELTSDHISMVIOKYLKSDQELSEESKONAHMLNIRMYLQNPAS
PMTVPPIYHSRNSKLVMPDIHCCYCDIVDDNLIDSVSEPTLIVHEDIMKRA
WIKVPCSLPRLIPENNGPFGSGRQREPLVIRINILIEVYSVDIRKELLOMNDANA
TECFMIDMRNNDIRENLDPGMACHGALNSFNNSSESDSDFLNTIRGLSLKRG
EYKVGKFGELGNSVYHITDIPIMSEEFIMVDPNINHISIKIKDSNPGITINMSK
QOKRLKRFNPGKFPIDVFCGQPLAVEAPYVNGTLFRISFTQOEAKEVSSYCY
NTADIVSLVDEFSICGRLIIFQSVSMYKLYLKEETNPLAODYIIRKRVCSK
ALNAPVSVLKEAKLMKTKSSNKLPTQVPSCSIIQITVEFHHVFRADLOSP
LFRGPDPPATLFEEMAKSGOSKRPDELPKQTCQYTWLICTGMDEGALKFLNSG
RLGLVPCGACVGLIHEGTOQKRTYVPHIEVCYLPRIKTKLPHINKEFVTSNR
KEIKWTDTKGRNMTTFEHRVIVAVYDQALVLDLAIAGELTYTYIAWPDLDLHD
DFSVICKEGYEDIALHGKGEKELTFVSGKMWVMKVNFRLLDSILDRKDYSAKFI
LKLYKTKGSKNLCAVELPSSVAKGFEAGCKQILLENTESEKQFSEVFPNIOETEA
ELRDPPLNFEVLNKELEDFSGILRVTCVPCSLGHPVLVPSRLIHEGKRVAKLETRD
GRPYGSTDYDLYNPIILIKYVGLMAKDDIILMDPMLEKRSVALEIKSHOAAACLRSS
ILSLIDEKTIKIDPRAKDPAKYOTIPIPLPTTKRAGSELKKGSKFRKTEYPAATD
IYAEYQDIDVCLLPILNENSHSFRCGGSVSLAVKELGLLKRPTVDLVINQLQYAK

SVNDCITLYQENITNACRYLHEAVLQNMAMATITIKLPCFCILVENYVESEKVS
 FHLNFEAAYLYLQPNKRYKNNRELFESEVQSFVEFPAFLVLEIDDERCKQJTE
 ENFOLCRRIISGILSLREKROEFCENYVGLIPDLMLLPALSLCYNDPMKPYK
 KSDTVKCHADIPREVAAYVGLAIPKRKHLKEVASICATGALTEGOKELTSRKS
 ILNAYPSEKEMKELLONDDAKATEICFPPDRPHOPVDRITDDKAPLOGALCYVN
 NOPFTEDDYRGIONIGKCKRECNPKCTGHYIGFNSVYHITTCPPSTISNDLIGTIDP
 HARVAPATVSPGNAFRDLDADPFOESVDVLDLYGNFKLDNCTMCFPLRNAMA
 OVSELSVSPSSDRMVOMLDLKLSDAEALMLLNHEKLSICEIDNCTMCFPLRNAMA
 GKITDGRKLRKQFHASVYDSYTKRQLKDIPVOQITTYMDRESCGNLTVMICRS
 GFSMEKYSKVSIAHKNDITLPPRGGAACITNNYKCKPHAFCSPLTLEGLDFPH
 VNGHFLDASRRRLMRDNGVGRSDMNSMTALAPAYVLLDOKRRTYPPGSPPT
 LSVLQNTLPHVKTPLKFLSRFPVNRDLQPDCLVALNLSCHIEDKRLPLPVRA
 PNIDSDLSHSAVITWINSTNKTREPFVDNLQDELQHLKADVNIITTRKVAENVY
 RLKHLLEIGFNLVNCDETAMLYHOLVDADIPVSPADVRSFLMTSPSSPDNCHI
 GKLPCRLQOTNLKLFHSLKLDVDFCDESEEFVEGGLPLITLTDVQIDPGRKPK
 FLTTYHELIPSRKDLPMNTLYKTSVYLNCVAKVDFIDSSPADLSSVLPREYKKN
 CAKKMDNFASBSWMLKNAHFISESYSVDDQDEPRAPVYDIDILKDALILGRTYV
 STSOLVPEBGDVLPLSLMHLAVFPAQSDKVFHALMGCIOLALNKICSDASLVP
 LLSCHTANIDSPASILKAVYVQSTFETKELMENDFALLMYFPCNLSHMSODDI
 KILSKPCYKSIISGRMSIAKFGTCVYLKSIPISEAWEKQSSSAFLEEKVLAKEI
 YVLGCVPDLEVLKHLPLKILENLSYAKLEHLIYLNKRLASIEPEIKLEOLPEK
 LESLIITHDANRLKOAKHFYDRTVRFVEVMLPEKFLIPKEFKLEOYIKRNOAF
 MTSWYEFLLNIGLKAYALSOOOLQRAKETSANENMSEKELQSVLDLHHITRER
 MDLSGNFLKELSLIPFCPEAPAEYIRFHPQOENGTPLKFNQAQVNPKEKQC
 DVQLQMLTSCPLPEKAPPLSTKEQEGSOLAQEOLEQVLMNLVMDPLKVINNC
 RNICHTITLDEMVKTRAKVLRISYIEFSAERREFFQALGVAFVAVEDEGWLKREE
 VYINLEYEADFPYLYKLPLELGTGHQLEKHLGTEIDISTKQYEVLESRKISSECKO
 LDPMERTRYKRVYSGLFKSDMSVNRSDLNNARLALYLPDSODKLVKSSILVDD
 APHYKSRIOGNIGVOMLVDSOCYIGKDGHPKTLIMEFORKLRKELSSILEODE
 ETPKVCORALCSLOGRLOLLSREQFITGLIRIKKHENDNAFLANEKATRICAKLR
 ECKLASCSEKLOTLTRKVGFNIPHSRSTETPAFLKFGNAVILLIYIHSDSNDINLL
 ALAMTLKSAITDMLISDTSTYLMALMCNDYRISSEKLDAGVYDSSEPKLEIPMGCT
 PIPAIHITLMDPMNVFPGYVGLVDAEGDLYGSQPTYVAIYQVEEREDAD
 NPSFGIKYQIDIGYSEKIVSSIDLKFSRDESSONDSAPTPPTSPTELEPDLR
 STPLPESGSEHSKSPSKHSPKRLKVNALPPLIKEVTSVQAKMLPSEKKIIR
 LYLKHPDNPNPHDIANEVFRHLDQELNRLKQAFLLDNAPRASRITSTASRQUS
 DKYSFQRTYSWNOBETSHKSRQOQKCPSPSAGQTYSPFPRKSYGNPEA
 RMLROARANFSAARNDLKNNANENWCFKYSTLKALIALAAYAVAGKSDKVPKPAL
 AOKIEYSQOLEGLTNDVHTLEBAYGVDSLKTRYPDLLPPOQIPNDRTSEVAMRVMEC
 TACIIIKLENFTQOKY"

BASE COUNT 3599 a 2281 c 2387 g 3226 t
 ORIGIN

Query Match 7.0%; Score 35; DB 10; Length 11493;
 Best Local Similarity 100.0%; Pred. No. 1.5e-07;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 203 gattgacttattatgagatgcactatccc 237
 Db 127 GATTGACTTATTGATGAGATGCACCTATCCC 161

RESULT 8
 LOCUS U29187 39760 bp DNA linear ROD 18-JAN-2000
 DEFINITION Mus musculus long incubation prion protein (Prnpb) and prion-like
 protein (Prnd) genes, complete cds.
 ACCESSION U29187
 VERSION U29187
 KEYWORDS U29187.1 GI:5281065
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 39760)
 Lee,I.Y., Westaway,D., Smit,A.F.A., Wang,K., Seto,J., Chen,L.,
 Acharya,C., Ankener,M., Baskin,D., Cooper,C., Yao,H., Prusiner,S.B.
 and Hood,L.E.
 TITLE Complete genomic sequence and analysis of the prion protein gene
 region from three mammalian species
 JOURNAL Genome Res. 8 (10), 1022-1037 (1998)
 MEDLINE 99018115

PUBMED 9799790
 REFERENCE 2 (bases 1 to 39760)
 AUTHORS Moore,R.C., Lee,I.Y., Silverman,G.L., Harrison,P.M., Strome,R.,
 Heinrich,C., Karunaratne,A., Pasternak,S.H., Chishti,M.A.,
 Liang,Y., Mastrangelo,P., Wang,K., Smit,A.F.A., Katamine,S.,
 Carlson,G., Cohen,F.E., Prusiner,S.B., Melton,D.W., Tremblay,P.,
 Hood,L.E. and Westaway,D.
 TITLE Ataxia in prion protein (Prp)-deficient mice is associated with
 upregulation of the novel Prp-like protein doppel
 J. Mol. Biol. 292 (4), 797-817 (1999)

JOURNAL 99457485
 MEDLINE 10525406
 PUBMED 3 (bases 1 to 39760)
 REFERENCE Lee,I.Y.
 TITLE Direct Submission
 JOURNAL Submitted (14-JUN-1995) Department of Molecular Biotechnology,
 University of Washington, Box 352145, Seattle, Washington
 98195-2145, USA
 4 (bases 1 to 39760)
 REFERENCE Lee,I.Y.
 TITLE Direct Submission
 JOURNAL Submitted (31-AUG-1999) Department of Molecular Biotechnology,
 University of Washington, Box 352145, Seattle, Washington
 98195-2145, USA

REMARK
 COMMENT Sequence update by submitter
 INTERPRETED Repeats were identified with RepeatMasker (available
 from <http://ftp.genome.washington.edu/RM/RepeatMasker.html>).
 FEATURES
 source 1..39760
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /chromosome="2"
 /clone="cos6.1/LnJ"
 /note="strain=1/LnJ"
 complement(1..167)
 /rpt_family="ORR1B"
 complement(1258..1677)
 /rpt_family="MER67C"
 1751..1863
 /rpt_family="B4"
 1857..1925
 /rpt_family="CA)n"
 1874..1925
 /rpt_unlt="(AC)x25
 2184..2500
 /rpt_family="ORR1A3"
 complement(2866..3091)
 /rpt_family="B3"
 4902..4940
 /rpt_unlt="(CA)x19
 complement(5239..5288)
 /rpt_family="B4"
 complement(5241..5326)
 /rpt_family="RSINE1"
 complement(5371..5422)
 /rpt_family="PB1D10"
 5572..5712
 /rpt_family="B1_MW"
 5572..5561
 /rpt_family="PB1D7"
 6205..6251
 /gene="Prnpb"
 /number=1
 join(6205..6251,8442..8539,19667..21675)
 mRNA
 gene
 repeat_region
 repeat_region
 repeat_region
 repeat_region
 exon


```

exon                /rpt_family="PB1D7"
                    8442..8539
                    /gene="Prnpb"
                    /number=2
repeat_region       /rpt_family="PB1D10"
                    complement(8894..8945)
repeat_region       /rpt_family="PB1D10"
                    complement(10249..10341)
repeat_region       /rpt_family="ID_RN"
                    complement(10269..10341)
repeat_region       /rpt_family="ID4"
                    complement(10598..10808)
repeat_region       /rpt_family="B2"
                    11198..11343
repeat_region       /rpt_family="B1_MM"
                    complement(11573..11651)
repeat_region       /rpt_family="ID_RN"
                    complement(11592..11651)
repeat_region       /rpt_family="ID5"
                    complement(12621..13162)
repeat_region       /rpt_family="L1MA4"
                    complement(13163..13262)
repeat_region       /rpt_family="PB1D10"
                    13809..14171
repeat_region       /rpt_family="ORR1D"
                    complement(14242..14376)
repeat_region       /rpt_family="B1_MM"
                    complement(15937..15985)
repeat_region       /rpt_family="ID6"
                    16797..16879
repeat_region       /rpt_family="LTR10A"
                    complement(17191..17372)
repeat_region       /rpt_family="LINE2"
                    17438..17462
repeat_region       /rpt_unit="(GATY)x6"
                    complement(17460..17526)
repeat_region       /rpt_family="ID4"
                    complement(17980..18149)
repeat_region       /rpt_family="L1_MM"
                    18517..18612
exon                /rpt_family="PB1D7"
                    19667..21675
                    /gene="Prnpb"
                    /number=3
                    19677..20441
                    /gene="Prnpb"
                    /codon_start=1
                    /product="long incubation prion protein"
                    /protein_id="A0D41440.1"
                    /db_xref="GI:5281066"
                    /translation="MANLGYWLALPVTMTDYGLCCKRRKPGGKMTGSGSRYPGQSGP
                    GGNRYPPQGGTQCPHGGWGQPHGGSWGQPHGGWGQGGGQGGGTHNDNKKSS
                    KPTNFKHVAAGAAGAVGGLGYMGASMRMIFGNDWEDRYRENNRYRPNQV
                    YRRVDQYQNQNNFVHDCVNITIKQHTVYTTGENTETDVKMKMERVHQQMCVTOYO
                    KESQAYYDGRSSSTVLFSSPYILLISFLIFLIVG"
repeat_region       complement(21935..22120)
                    /rpt_family="B3"
repeat_region       complement(22278..22387)
                    /rpt_family="B1_F"
repeat_region       complement(22419..22548)
                    /rpt_family="B1_F"
repeat_region       23599..23872
                    /rpt_family="B4"
repeat_region       complement(24414..24565)
                    /rpt_family="MER3A"
exon                25842..26004
                    /note="intergene exon 1"
repeat_region       complement(27124..27269)
                    /rpt_family="B1_MM"
repeat_region       complement(27289..27350)
                    /rpt_family="B1_MM"
repeat_region       27385..27520
                    /rpt_family="B1_MM"
repeat_region       27648..27833

```

```

repeat_region       /rpt_family="MLT2D"
                    27819..27910
repeat_region       /rpt_family="(GGA)n"
                    27828..27885
repeat_region       /rpt_unit="(GGA)x19"
                    28003..28120
repeat_region       /rpt_family="MLT2CB"
                    complement(28530..28725)
repeat_region       /rpt_family="MLT1B"
                    29572..29672
exon                /note="intergene exon 2"
                    30254..30345
repeat_region       /rpt_family="L1ME3"
                    30341..30573
repeat_region       /rpt_family="L1ME3A"
                    34086..39315
gene               /gene="Prnd"
                    /note="prion-like gene"
                    34086..34124
exon               /gene="Prnd"
                    /number=1
                    join(34086..34124,36205..37472,38013..39315)
                    /gene="Prnd"
                    /product="prion-like protein"
                    35130..35341
repeat_region       /rpt_family="UR1A"
                    36205..36799
exon               /gene="Prnd"
                    /note="exon 2a: short form of exon 2"
                    36205..37472
                    /gene="Prnd"
                    /note="exon 2b: long form of exon 2"
                    36212..36751
CDS                /gene="Prnd"
                    /codon_start=1
                    /product="prion-like protein"

Query Match      4.48; Score 22; DB 10; Length 39760;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      26  cagctgcagcttcaaatcag 47
Db      25620  CAGCTGCAGCTTCTAATCCAG 25641

RESULT      9
LOCUS       AE006329                10334 bp    DNA     linear    BCT 14-MAY-2001
DEFINITION  Lactococcus lactis subsp. lactis IL1403 section 91 of 218 of the
complete genome.
ACCESSION   AE006329 AE005176
VERSION     AE006329.1 GI:12723880
KEYWORDS
SOURCE      Lactococcus lactis subsp. lactis.
ORGANISM    Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
            Lactococcus.
REFERENCE   1 (bases 1 to 10334)
AUTHORS    Bojotin,A., Wincker,P., Manger,S., Jallion,O., Malarme,K.,
            Weissenbach,J., Ehrlich,S.D. and Sorokin,A.
TITLE      The Complete Genome Sequence of the Lactic Acid Bacterium
            Lactococcus lactis ssp. lactis IL1403
JOURNAL     Genome Res. 11 (5), 731-753 (2001)
MEDLINE     21235186
PUBMED     11337471
REFERENCE   2 (bases 1 to 10334)
AUTHORS    Bojotin,A., Wincker,P., Manger,S., Jallion,O., Malarme,K.,
            Weissenbach,J., Ehrlich,S.D. and Sorokin,A.
TITLE      Direct Submission
JOURNAL     Submitted (09-JAN-2001) INRA, Genetique Microbienne, Domaine de
            Vilvert, Jouy en Josas 78352, France
FEATURES

```

```

source
1. .10334
/organism="Lactococcus lactis subsp. lactis"
/strain="IL1403"
/sub_species="lactis"
/db_xref="taxon:1360"
71. .1129
/gene="yjha"
/note="LI69897"
CDS
71. .1129
/gene="yjha"
/note="EVIDENCE BY GENEMARK BIO16.01 Unknown"
/codon_start=1
/transl_table=11
/product="UNKNOWN PROTEIN"
/protein_id="AAK05042.1"
/db_xref="GI:12723881"
/translation="MKNRLLKGILVSIILVPALLIGSVYIVYAQYHRLPNMKTOT
INKPOTPOVEICGTYIKMTFENIGVKYSPRSNPFMDGGDDVRAKSAVSAIDEDIR
LINKRPFNGINIOELDMOGDGRSQOVNEPOMWRAGLPDTISVTIONDSALDPVPIK
INKASGIMTYSKEIESSSTRYLPLETFENFKPELDLRAEVSITLPIKNSDKREVEN
THLSAIFIDOKIQRQLTLFDEMKKYASKDYICADYAHLAGAKHELTMMKEEF
PTSLTKMRVAAPLNAPTYSLDYADVKKPKMTFGIIDEFLVSNNIKDIKIQTION
OFSVDHPVYMDSLEK"
complement(1155. .2096)
/gene="yjhb"
/note="LI71002"
complement(1155. .2096)
/gene="yjhb"
/note="EVIDENCE BY HOMOLOGY BIO15.01 HYPOTHETICAL.
General, there are three paralogs in L.lactis. 23%
identical to putative 1,4-dihydroxy-2-naphthoate
octaprenyltransferase of B.subtilis."
/codon_start=1
/transl_table=11
/product="HYPOTHETICAL PROTEIN"
/protein_id="AAK05043.1"
/db_xref="GI:12723882"
/translation="WTKTFPNTFRIOQLPALSLPAGCLFAFYEGSFHFPLLTF
FIGLASINLFVSAMNNLMDOXKALDPEYKKNNIIAVKSINFKLNLNCIALLDVVY
GLIVAFLENILAIPGICGLFIATFYTPGPASFRRPLGELIACIGSGFGLAVYI
NSYDGEFINEDLRNMETWTWDKRIILPILLVLGMCFGVONFMPSDNCDLEODIR
NERVLYPYLIKPLISKVFGLYLFAGCIILAIALGILPMSILNSTITALKYIPNIK
KRRAQGSSEFFHIQIDNMLFNASTALTLLGITRK"
2250. .2513
/gene="yjhc"
/note="LI72073"
2250. .2513
/gene="yjhc"
/note="EVIDENCE BY HOMOLOGY BIO15.02 HYPOTHETICAL.
Conserved. 48% identical to Mj1558 of Methanococcus
jannaschii. Probably FOF1-type ATP synthase alpha subunit,
evidenced by COGNITOR."
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK05044.1"
/db_xref="GI:12723883"
/translation="MRAWVTVVGADKIGIVAGVTATLAELANITIEISQTLMSGAFPM
MVVETONSDFSQAQELSRRKGALGVNIHQNAIRNAMHKL"
2630. .3967
/gene="yjhd"
/note="LI72471"
2630. .3967
/gene="yjhd"
/note="EVIDENCE BY HOMOLOGY BIO15.01 HYPOTHETICAL.
Conserved. 59% identical to Mj1665 protein of
Methanococcus jannaschii."
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK05045.1"
/db_xref="GI:12723884"
/translation="MDIIONIKETIAMIEQNFDIRTITMGISLSDCIDADIDKRAEKTI

```

gene
4067.4546
/gene="yjhe"
/note="L173848"
4067.4546
/gene="yjhe"
/note="EVIDENCE BY GENEMARK BIO16.01 Unknown"
/codon_start=1
/transl_table=11
/product="UNKNOWN PROTEIN"
/protein_id="AAK05046.1"
/db_xref="gi:12723885"
/translation="MAKVNFRKDYVETGLETSPVADSLAGIRANEARYFNKKKFEYL
TYPAHEKDEVALPEKILIEERNLFSKILLEVAIYENDELWPEEYEDGMVNLILY
EKVGGEGKKPRKRAVGIKLSVGMETPEKEGFKFARORSKLAGEIRSGSYPIKEOWL"
4614.5309
/gene="yjhf"
/note="L1174407"
4614.5309
/gene="yjhf"
/note="EVIDENCE BY HOMOLOG Y BIO06.09 glycolysis"
/codon_start=1
/transl_table=11
/product="phosphoglycerate mutase (Ecs.4.2.1)"
/protein_id="AAK05047.1"
/db_xref="gi:12723886"
/translation="MVNIYLYVRHGKTMFNTIGRAQGSNDPTLPAGEOGIIIEGLGFK
EGGIFDAVYSDSGRTIOTMGFIILANDNGEIPYTDKRIKRCWFCSPGGGDELF
DGVLRITTEGNNHTDEPPDEVIAANAIEPVDAGAAPMEKISGIMGPFIDIKE
SEAKAKNIIVYSHSGMTIGTFVKLAHPFLPRPHLDNGSVTHLSFEDGKTTICKVGM
SRTEGSEKRLAK"
5306.6052
/gene="dacB"
/note="L175147"
5306.6052
/gene="dacB"
/note="EVIDENCE BY HOMOLOG BIO03.02 Murein sacculus and
peptidoglycan. In E.faecalis corresponding enzyme medates
vancomycin resistance"
/codon_start=1
/transl_table=11
/product="D-alanyl-D-alanine carboxypeptidase (EC
3.4.16.4)"
/protein_id="AAK05048.1"
/db_xref="gi:12723887"
/translation="MKRRITPSLIVLLIIGALGFIILINDKSNSTNSAKRTSIS
SKEVAKKDDIPAKOSMDIIVLSRKSPOKMDPLTAVNGIODVIAVOOPIAA
AQOISPAEHLISGRSVSYOGELYOSYQAMNOCGTVNSGNPISSEELIKKVVQYS
OPRSEHSEOTGLADMSINDLSNESQAKVOYAALAPYGFVLRPECKKATGIGED
WHPFRVGLGNKMTENHLTLEELKLIPA"
6066.6470
/gene="yjhh"
/note="L1175880"
6066.6470
/gene="yjhh"
/note="EVIDENCE BY GENEMARK BIO16.01 Unknown"
/codon_start=1
/transl_table=11
/product="UNKNOWN PROTEIN"
/protein_id="AAK05049.1"
/db_xref="gi:12723888"
/translation="MLNRRRLLYVNVALFIYSSEMIKYLVHNNFSSQFNNLSTN
IKTLITKTSITSMOSILYLLIFALLIFLKRMRPTFILLETISALLIIDVAIIIFI
QHEVYGSLEPIAVAVITVYMNIMOTSNDNL"
6691.7734
/gene="hrca"
/note="L0212"

CDS	6691.. 7734 /gene="hrcA" /note="EVIDENCE BY HOMOLOG Y BIO1.01 Adaptations and atypical conditions. NEGATIVE REGULATOR OF CLASS I HEAT SHOCK GENES (GRPE-DNAK-DNAU AND GROELS OPERONS) . PREVENTS HEAT-SHOCK INDUCTION OF THESE OPERONS" /codon_start=1 /transl_table=11 /product="heat-inducible transcription repressor HrcA" /protein_id="AAO5050.1" /db_xref="GI:12723889"
gene	/translation="MTERORQILNLIVSLYAKHPPIGSKLIDLSIOASSATTIRNDMLFKTAKSSGLTGTSFVNLAPQRQQLVFEMWLDHSVLSTLTGTGEVRITQ FILPRTMTADLAFLSNLVKRLVGRVLDIHYMLRTEIPQIVORFYRVSEVQLQE SIFDPLFERLTVAGRNINFDYARDNLAELYLKEISDDERMLHEIREITINDENAVAF DNDERKYNKLIIISOKFEVIPRGSGTLVGVPMEDYORFLSVLDVAKVLTMLDSY VRYLDGNHYESK" 7764.. 8303 /gene="grpe" /note="L0273" 7764.. 8303 /gene="grpe" /note="Evidence by Homology Bio1.01 Adaptations and atypical conditions. STIMULATES, JOINTLY WITH DNAU, THE ATPASE ACTIVITY OF DNAK. HELPS TO RELEASE ADP FROM DNAK THUS ALLOWING DNAK TO RECYCLE MORE EFFICIENTLY" /codon_start=1 /transl_table=11
CDS	Query Match 4.2%; Score 21; DB 1; Length 10334; Best Local Similarity 100.0%; Pred. No. 4.8; Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy	206 ttgacttatttgatgatgt 226 Db 776 TTGACTTTATTGGATGAGATG 796
RESULT 10	AF030694 47573 bp DNA linear INV 03-NOV-2000
LOCUS	Plasmidium falciparum strain Dd2 heat shock protein 86 (HSP86), O1 (o1), O3 (o3), O2 (o2), CG8 (cg8), CG4 (cg4), CG3 (cg3), putative chloroquine resistance transporter (crt), CG9 (cg9), CG1 (cg1), CG6 (cg6), CG2 (cg2), and CG7 (cg7) genes, complete cds.
ACCESSION	AF030694
KEYWORDS	AF030694.2 GI:6724279
SOURCE	Malaria parasite P. falciparum.
ORGANISM	Plasmidium falciparum
REFERENCE	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
AUTHORS	1 (bases 1 to 47573) Su,X., Kirkman,L.A., Fujicka,H. and Wellens,T.E.
TITLE	Complex polymorphisms in an approximately kDa protein are linked to chloroquine-resistant P. falciparum in Southeast Asia and Africa
JOURNAL	Cell 91 (5), 593-603 (1997)
MEDLINE	98054002
REFERENCE	2 (bases 1 to 47573) Fidock,D.A., Nomura,T., Talley,A.K., Cooper,R.A., Dzekunov,S.M., Ferdig,M.T., Ursos,L.M.B., Bir Singh Sidhu,A., Naude,B., Deltsch,K.W., Su,X.Z., Wootton,J.C., Roepe,P.D. and Wellens,T.E. Mutations in the P. falciparum digestive vacuole transmembrane protein PfCPT and evidence for their role in chloroquine resistance MoJ. Cell 6 (4), 861-871 (2000)
JOURNAL	21000495
MEDLINE	3 (bases 1 to 47573) Su,X.-Z., Kirkman,L.A. and Wellens,T.E.
REFERENCE	Direct Submission
AUTHORS	Submitted (21-OCT-1997) National Institutes of Allergy and Infectious Diseases, National Institutes of Health, NIH Campus, Bethesda, MD 20892-0425, USA
JOURNAL	4 (bases 1 to 47573)
REFERENCE	
AUTHORS	Fidock,D.A.; Nomura,T.; Su,X.-Z.; Wootton,J.C. and Wellens,T.E.
TITLE	Direct Submission
JOURNAL	Submitted (21-JAN-2000) LPD, NIAID, Bldg 4, Room 126, Bethesda, MD 20892-0425, USA
REMARK	Sequence update by submitter
COMMENT	On Jan 21, 2000 this sequence version replaced g1:2828827.
FEATURES	Location/Qualifiers
SOURCE	1.. 47573
organism="Plasmodium falciparum"	
/strain="Dd2"	
/db_xref="taxon:5833"	
/note="Indochina"	
1849.. 4909	
/gene="HSP86"	
join(1849.. 3279, 4103.. 4909)	
/gene="HSP86"	
/codon_start=1	
/evidence="experimental	
/product="heat shock protein 86"	
/protein_id="AAC47837.1"	
/db_xref="GI:2642495"	
/translation="MSYTFPAFNADIROLMSLIINFYSKNKEIFLRLLINASDADLK IRYSITDROKLSAEPEFFRIIPDKNNNTLTIEDSGIGTKMDLNINLTGITRK APEALIASDISMITGPGFGFSAYLVADVIVYISKNNDEQYVESAGGSFTYTK DETNEKGKGTKITLIKEDOLELEKERIKVDLKXKSEFTSPFIKYCRONEKETT ASDEEEEGEGERGEDEEEKKKKTGDMKADSKEDEDEEKEDNDONTDHK VEDYTELENAEKKKKKKKKKKITTYGEMDELKOKPLMMRRPEETVEASAFYS LTNDMEDHLNVKHVSYGOLFEPKALFLTPRAPDEMDEKNNIKNIKLCIDMFELAERK CEETIEPLMLNFHVGVSDGPDLNISRESIQOKKIILVIKNIKLIKCIDMFELAERK ENKRFYEQFSKNIUKLICIHEDNANKRITITELRPOTSXSDEMIKLEYDRKENOK DIYYITGESINAVNSPFLEALKTKGFEVITYWDPIDEVVQOLKDGDVKLCCTE GLDIDDEEARKDFETLKAEGICRVIKDVLEKERVKVVVGQRTDSPCLVTSEFG WSAMERIMKAQALRDNSMTSYMSKKIMEINARHP1ISA.LKOKADKDKDYVKDLI MLFDTSLTSGFALPEPTTFSKRHMKILGSLIDEENNIDLPLEDTVATOSK MEED"	
2447.. 2612	
/rpt_family="microsatellite HSP86"	
/rpt_unit=gaa	
5255.. 5395	
/rpt_family="microsatellite BSM77"	
5743.. 5900	
/rpt_family="microsatellite BSM97"	
complement(5852.. 8182)	
/gene="O1"	
complement(5852.. 8182)	
/gene="O1"	
/codon_start=1	
/evidence="not_experimental	
/product="O1"	
/protein_id="AAC47848.1"	
/db_xref="GI:2642506"	
/translation="MMKDMYINISKKNNNNKSTNSYSKNNISIKYDNKKYPSPSM DSNNNSLNKRNGISHYRKGRSHSSNSNYDHYDKRRNNIRNRSTYSSSSLNE EKKKKRRRSNIIIGDSENSIRKNSNSTKSSNSTSRNSNNNNHKRSRY DSYDNSYSYRSKIKSKYKDYKHQSRYNNNNNNNNNNRRFRYHDDDLNNKOKKH SDHIHNNOIDIDIKSYKLVNNTIKERIDKRYHRVGYGTHESANNYNNLOKMDH YRSITPDENKIRPKSSLSPTMENVENSISEDEKERYKSNSKHHDKKKKKERN HRSSDSSESISTCSDEKSSNKSSRSYEVKAKODDKKKKREEKKHKKKKHKK HNEKRKSKDKDYKDYKDYKDYKSKDYKDYKDYKDYKDYKDYKDYKDYKDYKSK NKSODESKDESKSHRRKARRKRTINOKIESIKKKRRKARRKTIINIKNNLIUT IEIKONKESDESNESSSTARSISFLCGYSDLNKKYLINNHKIKLEDEEEKLEVH NNKKYLKINNHHLSDIOSNKESDKYKERRKIKERKIKREKIDDEEYTNNDIK NYNNRPDELINDSNSDNESIGPKPLDVNYVLAKROIYGVAMPBGGAIAOFV OKCKRTRRREVGLSAAIENPESLGIVMGSRKRRANAIRMKENOVYSAEQRALA MYNTERANRENALINDLKEIKRONNSILSNEKNE"	
6772.. 7042	
/rpt_family="microsatellite p4"	
/rpt_unit=gkattbdtyctt	
complement(8139.. 8597)	
/gene="O3"	
complement(8139.. 8597)	
/gene="O3"	
/codon_start=1	
repeat_region	
repeat_region	
repeat_region	
gene	
CDS	
repeat_region	
gene	
CDS	


```

repeat_region /note="AluX repeat: matches 1. .304 of consensus
              AluX repeat: matches 1. .304 of consensus"
              32901. 32914
repeat_region /note="Charliel repeat: matches 2488. .2498 of consensus"
              32906. 32950
repeat_region /note="Charliel repeat: matches 1822. .1868 of consensus
              Charliel repeat: matches 1822. .1868 of consensus"
              34137. 34372
repeat_region /note="LIME repeat: matches 5610. .5858 of consensus
              LIME3A repeat: matches 5610. .5858 of consensus"
              35046. 35354
repeat_region /note="AluX repeat: matches 1. .308 of consensus
              AluX repeat: matches 1. .308 of consensus"
              37800. 37879
repeat_region /note="8 copies 10 mer cccacacac 71% conserved"

Query Match 4.2%; Score 21; DB 9; Length 130192;
Best Local Similarity 100.0%; Pred. No. 4.6;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 170 ttgtgaaatactttatata 190
Db 76089 GTTGGAAAAATCTTATATA 76069

RESULT 12
AC099718/c 163721 bp DNA linear HTG 18-NOV-2001
LOCUS Mus musculus clone RP23-368E4, WORKING DRAFT SEQUENCE, 6 unordered
DEFINITION pieces.
ACCESSION AC099718 GI:16974217
VERSION AC099718.1
KEYWORDS HTGS PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 163721)
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Mus musculus, clone RP23-368E4
Unpublished
2 (bases 1 to 163721)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhalter, B.,
Brown, A., Camarata, J., Campiano, A., Chang, J., Chazaro, B.,
Chopel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,
Cooke, P., DeRellano, K., Dewar, K., Diaz, D., Dodge, S., Faro, S.,
Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Heatford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
Jones, C., Kamat, A., Karatas, A., Kells, C., Larocque, K.,
Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,
Maclean, C., MacDonald, P., Major, J., Margulis, N., Matthews, C.,
McCarthy, M., McEwan, P., McPherson, K., McPherson, R., Meldrum, J.,
Meneus, L., Minova, T., Mlenka, V., Murphy, T., Naylor, J., Nguyen, C.,
Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,
Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Risse, C., Rogov, P.,
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R.,
Seaman, S., Severy, P., Spencer, B., Stange, Thoman, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Triggilo, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W., J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (18-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR

```

```

Web site: http://www.seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L17683
Center clone name: 368_E_4
----- Summary Statistics
Sequencing vector: Plasmid; N/A; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 161403 bases at least Q40
Consensus quality: 162593 bases at least Q30
Consensus quality: 162938 bases at least Q20
Insert size: 163000; agarose-fp
Insert size: 163221; sum-of-ctrls
Quality coverage: 8.7 in Q20 bases; agarose-fp
Quality coverage: 8.7 in Q20 bases; sum-of-ctrls
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 1925: contig of 1925 bp in length
* 1926 2025: gap of 100 bp
* 2026 2677: contig of 652 bp in length
* 2678 2777: gap of 100 bp
* 2778 3978: contig of 1201 bp in length
* 3979 4078: gap of 100 bp
* 4079 25935: contig of 21857 bp in length
* 25936 26035: gap of 100 bp
* 26036 132236: contig of 106201 bp in length
* 132237 132336: gap of 100 bp
* 132337 163721: contig of 31385 bp in length.
Location/Qualifiers
1. 163721
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP23-368E4"
/clone_lib="RP23-368E4 Female Mouse BAC"
1. 1925
/note="assembly-fragment"
clone_end:sp6
vector_side:left"
misc_feature 2026..2677
/note="assembly-fragment"
2778..3978
/note="assembly-fragment"
4079..25935
/note="assembly-fragment"
26036..132236
/note="assembly-fragment"
132337..163721
/note="assembly-fragment"
clone_end:r7
vector_side:right"
BASE COUNT 51477 a 30666 c 30186 g 50867 t 525 others
ORIGIN
Query Match 4.2%; Score 21; DB 2; Length 163721;
Best Local Similarity 100.0%; Pred. No. 4.5;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 375 ttgtcctaataattagatg 395
Db 150753 TTGTCTTAAAAATTTAGATG 150733

RESULT 13
AC018633/c

```

LOCUS AC018633 169234 bp DNA Linear PRI 21-JAN-2000
 DEFINITION Homo sapiens clone UMGc:djs1 or RP11-16G1 from 7p14-15, complete
 sequence.
 ACCESSION AC018633
 VERSION AC018633.2 GI:6729063
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 169234)
 Kaul,R.K., Yu,D., Wong,G.K.-S., Magness,C.L., Green,E.D., Green,P.
 and Olson,M.V.
 Large-scale MCD Mapping and Sequencing of Human Chromosome 7
 Unpublished
 2 (bases 1 to 169234)
 Bubb,K.L., Desmarais,C.L., Ramsey,S.A. and Hubley,R.M.
 Direct Submission
 Submitted (15-DEC-1999) Human Genome Center, University of
 Washington, Box 352145, Seattle, WA 98195, USA
 3 (bases 1 to 169234)
 Kaul,R.K. and Richards,B.K.
 Direct Submission
 Submitted (21-JAN-2000) Human Genome Center, University of
 Washington, Box 352145, Seattle, WA 98195, USA
 On Jan 21, 2000 this sequence version replaced gi:6579285.

COMMENT

----- Genome Center:

Center: University of Washington Genome Center

Center code: UMGc

Web site: http://genome.washington.edu

Contact: umgchelp@u.washington.edu

----- Project Information

Center project name: chr-7

Center clone name: djs1 (RP11-16G1)

----- Summary Statistics

Sequencing vector: M13; 100% of reads

Chemistry: Dye-primer Bodipy; 90% of reads

Chemistry: Dye-terminator Big Dye; 10% of reads

Assembly program: Phrap; version 0.990319

Insert size: 169 234; sum-of-ctrls

Quality coverage: 11.1X in Q20 bases; sum-of-ctrls

----- Overlapping Sequences:

5': UMGc:djs21

3': UMGc:djs29

----- Sequence Quality Assessment:

This entry has been annotated with sequence quality estimates computed by the Phrap assembly program.

All manually edited bases have been reduced to quality zero.

Quality levels above 40 are expected to have less than

1 error in 10,000 bp.

Base-by-base quality values are not generally visible from the

Genbank flat file format but are available as part

of this entry's ASN.1 file.

Double stranded (DS) coverage: 96.4%

DS or two chemistry coverage: 97.1%

Single stranded regions: 10

----- Sequence Validation:

This sequence has been validated by Multiple Complete Digest

fingerprinting. Comparison of the experimentally derived digest

fragments with sequence-predicted fragments is given below.

The electronically-digested sequence consists of both insert and

vector, in order to accurately represent the entire circular BAC.

Small fragments below a variable cutoff (approximately 400-600 bp)

are not resolved in the fingerprint and hence do not appear

in the table. There are no significant remaining discrepancies

between the experimental and predicted values. Uniquely ordered

fragments are separated by dashed lines.

BgIII

EcoRI

NotI

FP	Seq	FP	Seq	FP	Seq
2055.00	2061.00	8954.00	8687.00	11080.00	11109.00
5487.00	5270.00	7648.00	7660.00	4611.00	4562.00
1913.00	1893.00	7648.00	7556.00	11080.00	11183.00
1508.00	1491.00	838.00	801.00	1377.00	1369.00
12901.00	13322.00	4399.00	4303.00	2193.00	2193.00
1004.00	992.00	4140.00	4094.00	3735.00	3692.00
529.00	503.00	660.00	643.00	614.00	599.00
10358.00	10526.00	3053.00	3100.00	3918.00	3851.00
3805.00	3819.00	5075.00	5106.00	614.00	593.00
1674.00	1645.00	6909.00	6771.00	2570.00	2518.00
2228.00	2170.00	1911.00	1874.00	1997.00	1995.00
3051.00	2974.00	3635.00	3584.00	6199.00	6176.00
14433.00	14779.00	3975.00	3996.00	4913.00	5009.00
1508.00	1475.00	897.00	864.00	2193.00	2184.00
1865.00	1874.00	1410.00	1399.00	12585.00	12586.00
2228.00	2269.00	5716.00	5790.00	1819.00	1821.00
2055.00	2043.00	3306.00	3304.00	709.00	690.00
1347.00	1325.00	3975.00	3893.00	6528.00	6484.00
4492.00	4449.00	897.00	876.00	1944.00	1934.00
3805.00	3757.00	7648.00	7475.00	3127.00	3138.00
3051.00	3031.00	3635.00	3611.00	16492.00	16238.00
1674.00	1676.00	838.00	824.00	2842.00	2741.00
10030.00	10132.00	3053.00	2987.00	6528.00	6572.00
4492.00	4478.00	10778.00	10756.00	10111.00	9669.00
3278.00	3209.00	1250.00	1234.00	11366.00	11758.00
1347.00	1339.00	2135.00	2108.00	4059.00	3979.00
8691.00	8644.00	5075.00	4993.00	1997.00	1990.00
1347.00	1331.00	897.00	875.00	2570.00	2445.00
4350.00	4341.00	1103.00	1082.00	2570.00	2576.00
12901.00	12601.00	3635.00	3602.00	11366.00	11207.00
3278.00	3339.00	11685.00	11548.00	10111.00	10101.00
12901.00	13224.00	1410.00	1385.00	13982.00	13688.00
3051.00	3072.00	2819.00	2826.00	-----	-----
5487.00	5453.00	19500.00	19348.00	-----	-----
2076.00	2128.00	4140.00	4057.00	-----	-----
1535.00	1554.00	2745.00	2716.00	-----	-----

```

-----
3051.00 3063.00 1734.00 1710.00
4775.00 4741.00 3053.00 3104.00
10030.00 10021.00 8954.00 8789.00
-----
1147.00 1151.00
2285.00 2234.00
897.00 894.00
688.00 677.00
966.00 955.00
688.00 669.00
-----
FEATURES
source      Location/Qualifiers
1.169234
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7p14-15"
/clone="djs1 (RP11-1661)"
/cell_line="Male Blood"
/clone_lib="RPC-11 Human Male BAC Library"
392..494
/rpt_family="MER21"
complement(1075..1364)
/rpt_family="Alu"
complement(2210..2373)
/standard_name="SWS2626"
/feature="GenBank Accession Number: G12923"
repeat_region
2594..2876
/rpt_family="Alu"
complement(3252..3702)
/rpt_family="MLT1"
complement(4489..4790)
/rpt_family="Alu"
complement(5307..5594)
/rpt_family="Alu"
5847..6152
/rpt_family="Alu"
complement(6970..7082)
/standard_name="SWS1000"
/feature="GenBank Accession Number: G00158"
repeat_region
complement(7545..7634)
/rpt_family="Alu"
complement(8959..9186)
/rpt_family="Alu"
10137..10423
/rpt_family="Alu"
repeat_region

Query Match      4.2% Score 21; DB 9; Length 169234;
Best Local Similarity 100.0%; Pred. No. 4.5;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 177 aaaaattatatacatattt 197
|||||
Db 133656 AAAATCTTATATACATTTT 133636

RESULT 14
LOCUS AF034998 1295 bp mRNA linear INV 26-JAN-1999
DEFINITION Hyphantria cunea immune-related Hdd1 mRNA, complete cds.
ACCESSION AF034998
VERSION AF034998.1 GI:4090963
KEYWORDS fall webworm moth.
SOURCE ORGANISM Hyphantria cunea

```

```

REFERENCE
AUTHORS Shin,S.W., Park,S.-S., Park,D.S., Kim,M.G., Kim,S.C., Brey,P.T. and Park,H.Y.
TITLE Isolation and characterization of immune-related genes from the fall webworm, Hyphantria cunea, using PCR-based differential display and subtractive cloning
JOURNAL Insect Biochem. Mol. Biol. 28 (11), 827-837 (1998)
MEDLINE 99035790
REFERENCE 2 (bases 1 to 1295)
AUTHORS Shin,S.W., Park,S.-S. and Park,H.-Y.
TITLE Direct Submission
JOURNAL Submitted (19-NOV-1997) Insect Resources Lab., Korea Research Institute of Bioscience and Biotechnology, P.O. Box 115, Yuseong, Taejeon 305-600, Korea

FEATURES
source      Location/Qualifiers
1.1295
/organism="Hyphantria cunea"
/db_xref="taxon:39466"
6..968
/codon_start=1
/product="immune-related Hdd1"
/protein_id="AAD09279.1"
/db_xref="GI:4090964"
/translation="MYVLAIVGLVSTASAKDLTVGREGSLKRFQNTASPTIMROA
KVTENVSDREVITVOYAIYDRREDKGEAVITIEGNGHKNVTIELKSPAVRFDEI
RVYAKOEDDIQSPVDGGOONIQNPFQSDIOVNVKQVQYAKVDTIGSPVSTQIPN
GQDKVOHPVVIQGEDNQOVRPNRDIQGOETPTIQTPEELQSKNQNOSDIABE
QNKEMQAVIDGVFQNKPGESKPTIAEPKPIQVVAQHPGSGKLDNIDHVRVIGIDG
QKVQVSTENKVKQGOQEKSTIOVQOTNDISOQAKQVNVPIPLYKQ"

BASE COUNT 472 a 246 c 255 g 322 t
ORIGIN

Query Match      4.0% Score 20; DB 3; Length 1295;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 194 ttttcagagattgacctt 213
|||||
Db 286 TTTTCAGAGATTGACTTT 305

RESULT 15
LOCUS AC107995 57205 bp DNA linear HTG 24-JAN-2002
DEFINITION Homo sapiens chromosome 15 clone RP11-577014 map 15, LOW-PASS SEQUENCE SAMPLING.
ACCESSION AC107995
VERSION AC107995.1 GI:18308723
KEYWORDS HTG; HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 57205)
Homo sapiens chromosome 15, clone RP11-577014
Unpublished
2 (bases 1 to 57205)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B., Brown,A., Camarata,J., Campolano,A., Chang,J., Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Huine,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Labocque,K., Lamazares,R., Landers,T., Lehoczeky,J., Levine,R., Liu,G., Maclean,C., Macdonald,P., Major,J., Margulis,N., Matthews,C., McCarthy,M.,

```


TITLE
JOURNAL
COMMENT

McEwan, P., McKernan, K., Meldrum, J., Meneus, L., Mihova, T.,
Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,
Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,
Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C.,
Rettel, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J.,
Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupack, R., Seaman, S.,
Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliou, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, A., Ye, W. J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Submitted (24-JAN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smith, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: MIBR

Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu

Center project name: L24565

Center Clone name: 577_O_14

NOTE: This record contains 69 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 706: contig of 706 bp in length
* 707 806: gap of 100 bp in length
* 807 1528: contig of 722 bp in length
* 1529 1628: gap of 100 bp in length
* 1629 2319: contig of 691 bp in length
* 2320 2419: gap of 100 bp in length
* 2420 3128: contig of 709 bp in length
* 3129 3228: gap of 100 bp in length
* 3229 3949: contig of 721 bp in length
* 3950 4049: gap of 100 bp in length
* 4050 4784: contig of 735 bp in length
* 4785 4884: gap of 100 bp in length
* 4885 5614: contig of 730 bp in length
* 5615 5714: gap of 100 bp in length
* 5715 6425: contig of 711 bp in length
* 6426 6525: gap of 100 bp in length
* 6526 7261: contig of 736 bp in length
* 7262 7361: gap of 100 bp in length
* 7362 8097: contig of 736 bp in length
* 8098 8197: gap of 100 bp in length
* 8198 8918: contig of 721 bp in length
* 8919 9018: gap of 100 bp in length
* 9019 9757: contig of 739 bp in length
* 9758 9857: gap of 100 bp in length
* 9858 10593: contig of 736 bp in length
* 10594 10693: gap of 100 bp in length
* 10694 11441: contig of 748 bp in length
* 11442 11541: gap of 100 bp in length
* 11542 12280: contig of 739 bp in length
* 12281 12380: gap of 100 bp in length
* 12381 13110: contig of 730 bp in length
* 13111 13210: gap of 100 bp in length
* 13211 13922: contig of 712 bp in length
* 13923 14022: gap of 100 bp in length
* 14023 14727: contig of 705 bp in length
* 14728 14827: gap of 100 bp in length

14828 15553: contig of 726 bp in length
* 15554 15653: gap of 100 bp in length
* 15654 16377: contig of 723 bp in length
* 16377 16476: gap of 100 bp in length
* 16477 17185: contig of 709 bp in length
* 17186 17285: gap of 100 bp in length
* 17286 18035: contig of 750 bp in length
* 18036 18135: gap of 100 bp in length
* 18136 18872: contig of 737 bp in length
* 18873 18972: gap of 100 bp in length
* 18973 19721: contig of 749 bp in length
* 19722 19821: gap of 100 bp in length
* 19822 20542: contig of 721 bp in length
* 20543 20642: gap of 100 bp in length
* 20643 21361: contig of 719 bp in length
* 21362 21461: gap of 100 bp in length
* 21462 22213: contig of 752 bp in length
* 22214 22313: gap of 100 bp in length
* 22314 23036: contig of 723 bp in length
* 23037 23136: gap of 100 bp in length
* 23137 23286: contig of 725 bp in length
* 23287 23961: gap of 100 bp in length
* 23962 24713: contig of 752 bp in length
* 24714 24813: gap of 100 bp in length
* 24814 25557: contig of 744 bp in length
* 25558 25657: gap of 100 bp in length
* 25658 26405: contig of 748 bp in length
* 26406 26505: gap of 100 bp in length
* 26506 27257: contig of 752 bp in length
* 27258 27357: gap of 100 bp in length
* 27358 28094: contig of 737 bp in length
* 28095 28194: gap of 100 bp in length
* 28195 28929: contig of 735 bp in length
* 28930 29029: gap of 100 bp in length
* 29030 29749: contig of 720 bp in length
* 29750 29849: gap of 100 bp in length
* 29850 30595: contig of 746 bp in length
* 30596 30695: gap of 100 bp in length
* 30696 31408: contig of 713 bp in length
* 31409 31508: gap of 100 bp in length
* 31509 32221: contig of 713 bp in length
* 32222 32321: gap of 100 bp in length
* 32322 33304: contig of 723 bp in length
* 33045 33144: gap of 100 bp in length
* 33145 33893: contig of 749 bp in length
* 33894 33993: gap of 100 bp in length
* 33994 34717: contig of 724 bp in length
* 34718 34817: gap of 100 bp in length
* 34818 35549: contig of 732 bp in length
* 35550 35649: gap of 100 bp in length
* 35650 36353: contig of 704 bp in length
* 36354 36453: gap of 100 bp in length
* 36454 37182: contig of 729 bp in length
* 37183 37282: gap of 100 bp in length
* 37283 37998: contig of 716 bp in length
* 37999 38098: gap of 100 bp in length
* 38099 38849: contig of 751 bp in length
* 38850 38949: gap of 100 bp in length
* 38950 39699: contig of 750 bp in length
* 39700 39799: gap of 100 bp in length
* 39800 40533: contig of 734 bp in length
* 40534 40633: gap of 100 bp in length
* 40634 41360: contig of 727 bp in length
* 41361 41460: gap of 100 bp in length
* 41461 42209: contig of 749 bp in length
* 42210 42309: gap of 100 bp in length
* 42310 43061: contig of 752 bp in length
* 43062 43161: gap of 100 bp in length
* 43162 43892: contig of 731 bp in length
* 43893 43992: gap of 100 bp in length
* 43993 44724: contig of 732 bp in length
* 44725 44824: gap of 100 bp in length
* 44825 45545: contig of 721 bp in length

```
* 45546 45645: gap of 100 bp
* 45646 46393: contig of 748 bp in length
* 46394 46493: gap of 100 bp
* 46494 47186: contig of 693 bp in length
* 47187 47286: gap of 100 bp
* 47287 48011: contig of 725 bp in length
* 48012 48111: gap of 100 bp
* 48112 48835: contig of 724 bp in length
* 48836 48935: gap of 100 bp
* 48936 49667: contig of 732 bp in length
* 49668 49767: gap of 100 bp
* 49768 50519: contig of 752 bp in length
* 50520 50619: gap of 100 bp
* 50620 51350: contig of 731 bp in length
* 51351 51450: gap of 100 bp
* 51451 52179: contig of 729 bp in length
* 52180 52279: gap of 100 bp
* 52280 52997: contig of 718 bp in length
* 52998 53097: gap of 100 bp
* 53098 53825: contig of 728 bp in length
* 53826 53925: gap of 100 bp
* 53926 54648: contig of 723 bp in length
* 54649 54748: gap of 100 bp
* 54749 55501: contig of 753 bp in length
* 55502 55601: gap of 100 bp
* 55602 56350: contig of 749 bp in length
* 56351 56450: gap of 100 bp
* 56451 57205: contig of 755 bp in length.
```

Query Match 4.0%; Score 20; DB 2; Length 57205;
Best Local Similarity 100.0%; Pred. No. 16;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 28 gctgcagcttctaattccag 47
|||||

Db 6794 GCTGCAGCTTCTAATCCAG 6813

Search completed: May 22, 2002, 08:30:10
Job time: 10700 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 22, 2002, 06:46:20 ; Search time 373 Seconds
(without alignments)
2301.494 Million cell updates/sec

Title: US-09-693-205-7_COPY_1_500

Perfect score: 500
Sequence: 1 atgatttcaggaggaacatc.....gcagaatctgttaatacaaa 500

Scoring table:
OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1736436 seqs, 858457221 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

N_Geneseq_032802.*
1: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	500	100.0	12792	22 AAH20176	Human mutated spas
2	500	100.0	12793	22 AAH20174	Human spastin nucle
3	500	100.0	12793	22 AAH20178	Human mutated spas
4	500	100.0	12793	22 AAH20179	Human mutated spas
5	500	100.0	12793	22 AAH20182	Human mutated spas
6	35	7.0	11453	22 AAH20175	Mouse spastin nucle
7	23	4.6	23	22 AAH20130	Human spastin ORF
8	19	3.8	721	16 AAT35105	Down-regulated sen
9	19	3.8	1543	23 AAS83269	DNA encoding novel

10	19	3.8	1588	22 AAH72857	Human cervical can
C 11	19	3.8	3365	18 AAV74557	Staphylococcus aur
C 12	19	3.8	4057	23 ABL14512	Drosophila melanog
C 13	19	3.8	7076	24 ABL34145	Human immune syste
C 14	18	3.6	431	22 AAH12392	Human cDNA clone (
C 15	18	3.6	513	22 AAF55109	Nucleotide sequenc
C 16	18	3.6	516	22 AAF55108	Nucleotide sequenc
C 17	18	3.6	1034	23 AAS87201	DNA encoding novel
C 18	18	3.6	1440	23 AAH3781	S. epidermidis ope
C 19	18	3.6	1481	21 AAC47212	Arabidopsis thalia
C 20	18	3.6	1489	21 AAC39202	S. epidermidis gen
C 21	18	3.6	3180	22 AAH54577	Rat brain specific
C 22	18	3.6	3736	20 AAH94918	Human metastasis a
C 23	18	3.6	6012	24 ABL34554	Chemically pretrea
C 24	18	3.6	6203	22 AAS45475	Human immune syste
C 25	18	3.6	6731	24 ABL33061	Human nervous syst
C 26	18	3.6	9391	22 ABA16800	Genomic sequence #
C 27	18	3.6	10078	22 AAS40045	Human digestive sy
C 28	18	3.6	10078	22 AAK91462	Caenorhabditis ele
C 29	18	3.6	11459	20 AAX06876	Human immune syste
C 30	18	3.6	15667	24 ABL34146	BAC containing rep
C 31	18	3.6	67087	21 AAF22280	Genomic fragment #
C 32	18	3.6	100848	22 AAF28552	Human SNP oligonuc
C 33	17	3.4	51	22 AAL32440	Genomic sequence #
C 34	17	3.4	287	22 AAS28202	Genomic sequence #
C 35	17	3.4	287	22 AAS28204	CDNA sequence of p
C 36	17	3.4	305	19 AAV61286	Prostate tumour sp
C 37	17	3.4	305	19 AAV58543	Human immunogenic
C 38	17	3.4	305	17 AAO6306	Human prostate CDN
C 39	17	3.4	305	22 AAS63514	Human prostate tum
C 40	17	3.4	305	22 AAS10065	Human prostate-spe
C 41	17	3.4	305	22 AAH93422	Human prostate-spe
C 42	17	3.4	305	22 AAH84736	Prostate tumour an
C 43	17	3.4	305	22 AAH02487	Human immune/haema
C 44	17	3.4	334	22 AAK06614	Human polynucleoti
C 45	17	3.4	382	22 AAI92739	

ALIGNMENTS

RESULT 1	
AAH20176	AAH20176 standard; DNA; 12792 BP.
ID	AAH20176 standard; DNA; 12792 BP.
XX	AAH20176;
AC	09-AUG-2001 (first entry)
XX	
DE	Human mutated spastin nucleotide sequence SEQ ID NO:7.
XX	
KW	Human; mouse; spastin; ARSACS; chromosome 13q11; identification;
KW	autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;
KW	neurodegenerative disease; reduced sensory nerve conduction; diagnosis;
KW	reduced motor nerve velocity; hypermyelination of retinal nerve fibre;
KW	atrophy of upper cerebellar vermis; absence of Purkinje cell;
KW	abnormal neuronal lipid storage; genetic disorder; characterisation; ds.
XX	
OS	Homo sapiens.
XX	
XX	Synthetic.
XX	
FT	Key
FT	Location/Qualifiers
FT	77..6604
FT	/*tag="a
XX	/product="mutated spastin"
XX	
XX	W0200129266-A2.
PN	
XX	
XX	26-APR-2001.
PD	
XX	
XX	20-OCT-2000; 2000WO-US29130.
PF	
XX	
XX	20-OCT-1999; 99US-0160588.
PR	

XX (UYMC-) UNIV MCGILL.
PA (HOPI-) HOSPITAL SAINTE-JUSTINE.
XX
XX
PI Hudson TJ, Engert J, Richter A;
XX
XX WPl: 2001-308494/32.
DR P-PSDB; AAB97821.
XX
XX New isolated polynucleotide, encoding spastin gene, and polypeptides,
PT useful for diagnosing autosomal recessive spastic ataxia of
PT Charlevoix-Sequenay disease by detecting two point mutations in spastin
PT gene sequence -
XX
XX Claim 1; Page -: 76pp; English.

CC The present invention describes human and mouse spastin, and mutated
CC human spastin (autosomal recessive spastic ataxia of Charlevoix-Seguenay
CC (ARSACS)) gene sequences (1). The spastin gene has been mapped to
CC chromosome 13q11. (1) have neuroprotective activities and can be used in
CC gene therapy and as a spastin polypeptide agonists. (1), their fragments
CC or their complements can be useful for assaying the presence of a nucleic
CC acid molecule in a sample. (1) is useful for diagnosing or aiding in the
CC diagnosis of an early onset neurodegenerative disease in an individual.
CC The neurodegenerative disease comprises reduced sensory nerve conduction,
CC reduced motor nerve velocity, hypomyelination of retinal nerve fibres,
CC atrophy of upper cerebellar vermis, absence of Purkinje cells and
CC abnormal neuronal lipid storage. (1) can also be used to produce
CC antisense nucleic acids, is useful as molecular weight or chromosome
CC markers, to identify genetic disorders, as hybridisation probes or
CC primers, as an antigen, identify and express recombinant protein for
CC analysis, characterisation or therapeutic use, or as markers for tissues
CC in which the corresponding protein is expressed. Diagnostic methods from
CC the present invention can be used to identify subjects having or at risk
CC of developing a disease or disorder associated with aberrant expression
CC or activity of (1). The assays can be utilised to identify a subject
CC having or at risk of developing a disorder associated with Spastin
CC protein or spastin gene expression or activity. The present sequence
CC encodes a mutated human spastin.
CC N.B. The present sequence is not given in the present specification but
CC is derived from the human spastin nucleotide sequence (AAH20174) as
CC stated on page 14.

Sequence 12792 BP; 4163 A; 2256 C; 2487 G; 3886 T; 0 other;

Query Match	100.0%;	Score 500;	DB 22;	Length 12792;
Best Local Similarity	100.0%;	Pred. No. 4.7e-239;		
Matches 500;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY 1 atgatttaacgaagaagaccatctactcaagctcgcaagctcttaacccgaagaagcttttcag 60
Db 1 atgatttaacgaagaagaccatctactcaagctcgcaagctcttaacccgaagaagcttttcag 60
QY 61 tcttatcaagaagaatgaatgaatacatctcgcccgccgagagaatctgatcttcaatgtrta 120
Db 61 tcttatcaagaagaatgaatgaatacatctcgcccgccgagagaatctgatcttcaatgtrta 120
QY 121 tccatttgatgtaaaacagaaatccaccatctgtrtttcaatggcttaagaatggtttggaanaa 180
Db 121 tccatttgatgtaaaacagaaatccaccatctgtrtttcaatggcttaagaatggtttggaanaa 180
QY 181 tctttatatacatcttttccaagagatttgacttaatttgatgtagatgaccattatccccag 240
Db 181 tctttatatacatcttttccaagagatttgacttaatttgatgtagatgaccattatccccag 240
QY 241 aactatactagaagaagtcagaaatgtgtggaactcattatagactcaagaattccatcgtt 300
Db 241 aactatactagaagaagtcagaaatgtgtggaactcattatagactcaagaattccatcgtt 300
QY 301 agtcaattttagacatgaatcctctagaacagcgtctccagaattttttgacagaactgttaca 360
Db 301 agtcaattttagacatgaatcctctagaacagcgtctccagaattttttgacagaactgttaca 360

QY 361 aaacttggaggttttgccttaaaaattagatgacatcatatacaatcgttttta 420

Db 361 aaacttggaggttttgccttaaaaattagatgacatcatatacaatcgttttta 420

QY 421 aaatatcttattcacattaccgaatgctgtttgcataatggagaagatgcatt 480

Db 421 aaatatcttattcacattaccgaatgctgtttgcataatggagaagatgcatt 480

QY 481 gcaagaattgttatacaca 500

Db 481 gcaagaattgttatacaca 500

RESULT 2

AAH20174
ID AAH20174 standard; DNA; 12793 BP.

AC AAH20174;

DT 09-AUG-2001 (first entry)

DE Human spastin nucleotide sequence SEQ ID NO:1.

KW Human; mouse; spastin; ARSACS; chromosome 13q11; identification;
 KW autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;
 KW neurodegenerative disease; reduced sensory nerve conduction; diagnosis;
 KW reduced motor nerve velocity; hypermyelination of retinal nerve fibre;
 KW atrophy of upper cerebellar vermis; absence of Purkinje cell;
 KW abnormal neuronal lipid storage; genetic disorder; characterisation; ds
 XX
 XS Homo sapiens.

OS Homo sapiens.

FH	key	Location/Qualifiers
FT		77.11566
FT	CDS	/*tag= a
FT		/product= "spastin"

PN WO200129266-A2.

PD 26-APR-2001.

PF 20-OCT-2000; 2000WO-US29130.

PR 20-OCT-1999; 99US-0160588.

PA (UYMC-) UNIV MCGILL.

XX

XX

DR P-PSDB; AAB97819.

PT New isolated polynucleotide, encoding spastin gene, and polypeptides,
PT useful for diagnosing autosomal recessive spastic ataxia of
PT Charlevoix-Saguenay disease by detecting two point mutations in spastin
PT gene sequence -

PS Claim 1; Fig 9; 76pp; English.

CC The present invention describes human and mouse spinin and mutated
CC human spinin (autosomal recessive spastic ataxia of Charlevoix-Saguenay
CC (ARSACS)) gene sequences (1). The spinin gene has been mapped to
CC chromosome 13q11. (1) have neuroprotective activities and can be used in
CC gene therapy and as a spinin polypeptide agonists. (1), their fragments
CC or their complements can be useful for assaying the presence of a nuclear
CC acid molecule in a sample. (1) is useful for diagnosing or aiding in the
CC diagnosis of an early onset neurodegenerative disease in an individual.
CC The neurodegenerative disease comprises reduced sensory nerve conduction
CC reduced motor nerve velocity, hypomyelination of retinal nerve fibres,
CC atrophy of upper cerebellar vermis, absence of Purkinje cells and
CC abnormal neuronal lipid storage. (1) can also be used to produce

CC antisense nucleic acids, is useful as molecular weight or chromosome
 CC markers, to identify genetic disorders, as hybridisation probes or
 CC primers, as an antigen, identify and express recombinant protein for
 CC analysis, characterisation or therapeutic use, or as markers for tissues
 CC in which the corresponding protein is expressed. Diagnostic methods from
 CC the present invention can be used to identify subjects having or at risk
 CC of developing a disease or disorder associated with aberrant expression
 CC or activity of (I). The assays can be utilised to identify a subject
 CC having or at risk of developing a disorder associated with Spastin
 CC protein or spastin gene expression or activity. The present sequence
 CC encodes human spastin as given in the present invention.
 XX
 XX Sequence 12793 BP; 4163 A; 2255 C; 2487 G; 3887 T; 0 other;

Query Match 100.0%; Score 500; DB 22; Length 12793;
 Best Local Similarity 100.0%; Pred. No. 4.7e-239;
 Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgatttaacaggaagaccatgtactcagctgcagcttctaataccagaagattgcagc 60
 |||
 Db 1 atgatttaacaggaagaccatgtactcagctgcagcttctaataccagaagattgcagc 60
 QY 61 tcttatacaaggaagtaataatattctgcccgcgcagagaagaattgttcaatgta 120
 |||
 Db 61 tcttatacaaggaagtaataatattctgcccgcgcagagaagaattgttcaatgta 120
 QY 121 tccatttgatgaagaacagaataccaccatctgtttcattagcttaagatggttgaaaaa 180
 |||
 Db 121 tccatttgatgaagaacagaataccaccatctgtttcattagcttaagatggttgaaaaa 180
 QY 181 tctttataatcattttccagaagatttgactttatttgatgagatgccattatcccccag 240
 |||
 Db 181 tctttataatcattttccagaagatttgactttatttgatgagatgccattatcccccag 240
 QY 241 aactatactagagaagagtcagacatgtgtcgaactcattagactcaggaattccatcgtt 300
 |||
 Db 241 aactatactagagaagagtcagacatgtgtcgaactcattagactcaggaattccatcgtt 300
 QY 301 agtcaatttgacagatgaatctgagacagcttcagaatttttagcagaacattgtaca 360
 |||
 Db 301 agtcaatttgacagatgaatctgagacagcttcagaatttttagcagaacattgtaca 360
 QY 361 aaactctggaggggttgccttaaaaaatragatgcatctataaacaatccgcttataa 420
 |||
 Db 361 aaactctggaggggttgccttaaaaaatragatgcatctataaacaatccgcttataa 420
 QY 421 aaaaatataatcattaccattacaagtgctgttttcagataaagtgagaagatgcatt 480
 |||
 Db 421 aaaaatataatcattaccattacaagtgctgttttcagataaagtgagaagatgcatt 480
 QY 481 gcagaaatgtgtaatacaaa 500
 |||
 Db 481 gcagaaatgtgtaatacaaa 500

RESULT 3
 AAH20178
 ID AAH20178 standard; DNA: 12793 BP.

XX AAH20178;
 DT 09-AUG-2001 (first entry)

DE Human mutated spastin nucleotide sequence SEQ ID NO:11.

KM Human; mouse; spastin; ARSACS; chromosome 13q11; identification;
 KM autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;
 KM neurodegenerative disease; reduced sensory nerve conduction; diagnosis;
 KM reduced motor nerve velocity; hypermyelination of retinal nerve fibres;
 KM atrophy of upper cerebellar vermis; absence of Purkinje cell;
 KM abnormal neuronal lipid storage; genetic disorder; characterisation; ds.
 XX

OS Homo sapiens.
 OS Synthetic.
 XX
 PN WO200129266-A2.
 XX
 XX 26-APR-2001.
 PD
 XX 20-OCT-2000; 2000MO-US29130.
 PF
 XX 20-OCT-1999; 99US-0160588.
 PR
 XX (UYMC-) UNIV MCGILL.
 PA (HOPI-) HOPITAL SAINTE-JUSTINE.
 XX
 XX Hudson TJ, Richter A;
 PI WPI; 2001-308494/32.
 DR
 XX
 XX
 PT New isolated polynucleotide, encoding spastin gene, and polypeptides,
 PT useful for diagnosing autosomal recessive spastic ataxia of
 PT Charlevoix-Saguenay disease by detecting two point mutations in spastin
 PT gene sequence -
 PS
 XX Claim 1; Page -: 76pp; English.

CC The present invention describes human and mouse spastin, and mutated
 CC human spastin (autosomal recessive spastic ataxia of Charlevoix-Saguenay
 CC (ARSACS)) gene sequences (I). The spastin gene has been mapped to
 CC chromosome 13q11. (I) have neuroprotective activities and can be used in
 CC gene therapy and as a spastin polypeptide agonists. (II), their fragments
 CC or their complements can be useful for assaying the presence of a nucleic
 CC acid molecule in a sample. (I) is useful for diagnosing or aiding in the
 CC diagnosis of an early onset neurodegenerative disease in an individual.
 CC The neurodegenerative disease comprises reduced sensory nerve conduction,
 CC reduced motor nerve velocity, hypermyelination of retinal nerve fibres,
 CC atrophy of upper cerebellar vermis, absence of Purkinje cells and
 CC abnormal neuronal lipid storage. (I) can also be used to produce
 CC antisense nucleic acids, is useful as molecular weight or chromosome
 CC markers, to identify genetic disorders, as hybridisation probes or
 CC primers, as an antigen, identify and express recombinant protein for
 CC analysis, characterisation or therapeutic use, or as markers for tissues
 CC in which the corresponding protein is expressed. Diagnostic methods from
 CC the present invention can be used to identify subjects having or at risk
 CC of developing a disease or disorder associated with aberrant expression
 CC or activity of (I). The assays can be utilised to identify a subject
 CC having or at risk of developing a disorder associated with Spastin
 CC protein or spastin gene expression or activity. The present sequence
 CC represents a mutated human spastin gene from the present invention.
 CC N.B. The present sequence is not given in the present specification but
 CC is derived from the human spastin nucleotide sequence (AAH20174) as
 CC stated on page 14.
 CC
 XX
 XX Sequence 12793 BP; 4163 A; 2255 C; 2487 G; 3888 T; 0 other;

Query Match 100.0%; Score 500; DB 22; Length 12793;
 Best Local Similarity 100.0%; Pred. No. 4.7e-239;
 Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgatttaacaggaagaccatgtactcagctgcagcttctaataccagaagattgcagc 60
 |||
 Db 1 atgatttaacaggaagaccatgtactcagctgcagcttctaataccagaagattgcagc 60
 QY 61 tcttatacaaggaagtaataatattctgcccgcgcagagaagaattgttcaatgta 120
 |||
 Db 61 tcttatacaaggaagtaataatattctgcccgcgcagagaagaattgttcaatgta 120
 QY 121 tccatttgatgaagaacagaataccaccatctgtttcattagcttaagatggttgaaaaa 180
 |||
 Db 121 tccatttgatgaagaacagaataccaccatctgtttcattagcttaagatggttgaaaaa 180
 QY 181 tctttataatcattttccagaagatttgactttatttgatgagatgccattatcccccag 240
 |||

Query Match	Best Local Similarity	100.0%	Score 500;	DB 22;	Length 12793;
Matches	500;	Conservative	0;	Mismatches	0;
				Indels	Gaps
QY 1	atgatttcaggaagaacatgactgactgaagcttcaatccagaagattgacg	60			
Db 1	atgatttcaggaagaacatgactgactgaagcttcaatccagaagattgacg	60			
QY 61	tctttcagaaggaatgaatgatacttcggcctggcagaagaattgatgttcaatgta	120			
Db 61	tctttcagaaggaatgaatgatacttcggcctggcagaagaattgatgttcaatgta	120			
QY 121	tcattttgatgaaacagaaatccaccatctgttccatgcttaagatggtttggaaaa	180			
Db 121	tcattttgatgaaacagaaatccaccatctgttccatgcttaagatggtttggaaaa	180			
QY 181	tctttatatacatcttttcagaagatttgacttatttgatgagatgcacattaccocag	240			
Db 181	tctttatatacatcttttcagaagatttgacttatttgatgagatgcacattaccocag	240			
QY 241	aactatactagaaggaagtcagactgtgtggaactcattagaaccagatccatcgtt	300			
Db 241	aactatactagaaggaagtcagactgtgtggaactcattagaaccagatccatcgtt	300			
QY 301	agtcatttcagaacgtgatcctgaagcagagcttcacagaatttttagcagacatgtaca	360			
Db 301	agtcatttcagaacgtgatcctgaagcagagcttcacagaatttttagcagacatgtaca	360			
QY 361	aaaactgtgaggggtttgtctcttaaaaaatttagatgcattctatacaaatccgctattaa	420			
Db 361	aaaactgtgaggggtttgtctcttaaaaaatttagatgcattctatacaaatccgctattaa	420			
QY 421	aaaatatcttcatctaccattaccagaagtgctgttttgagaataagggagaagtgcatt	480			
Db 421	aaaatatcttcatctaccattaccagaagtgctgttttgagaataagggagaagtgcatt	480			
QY 481	gcagaagaattgtgtaatcaaa	500			
Db 481	gcagaagaattgtgtaatcaaa	500			
RESULT	5				
AAH20182					
ID	AAH20182	standard; DNA;	12793	BP.	
XX	AAH20182;				
XX	09-AUG-2001	(first entry)			
XX					
XX	Human mutated spastin nucleotide sequence SEQ ID NO:15.				

KM Human; mouse; spastin; ARSACS; chromosome 13q11; identification;
KM autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;
KM neurodegenerative disease; reduced sensory nerve conduction; diagnosis;
KM reduced motor nerve velocity; hypermyelination of retinal nerve fibres;
KM atrophy of upper cerebellar vermis; absence of Purkinje cell;
KM abnormal neuronal lipid storage; genetic disorder; characterisation; ds.
XX
OS Homo sapiens.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT CDS 77..11566
FT /*tag= a
FT /product= "mutated spastin"
XX
PN WO200129266-A2.
XX
PD 26-APR-2001.
XX
PF 20-OCT-2000; 2000MO-US29130.
XX
PR 20-OCT-1999; 99US-0160588.
XX
PA (UYMC-) UNIV MCGILL.
XX (HOP1-) HOPITAL SAINTE-JUSTINE.
PI Hudson TJ, Engert J, Richter A;
XX
DR WPI; 2001-308494/32.
XX P-PDB; AAB97823.
XX
PT New isolated polynucleotide, encoding spastin gene, and polypeptides,
PT useful for diagnosing autosomal recessive spastic ataxia of
PT Charlevoix-Saguenay disease by detecting two point mutations in spastin
PT gene sequence -
XX
PS Claim 1; Page -: 76pp; English.
XX
CC The present invention describes human and mouse spastin, and mutated
CC human spastin (autosomal recessive spastic ataxia of Charlevoix-Saguenay
CC (ARSACS)) gene sequences (I). The spastin gene has been mapped to
CC chromosome 13q11. (I) have neuroprotective activities and can be used in
CC gene therapy and as a spastin polypeptide agonists. (I), their fragments
CC or their complements can be useful for assaying the presence of a nucleic
CC acid molecule in a sample. (I) is useful for diagnosing or aiding in the
CC diagnosis of an early onset neurodegenerative disease in an individual.
CC The neurodegenerative disease comprises reduced sensory nerve conduction,
CC reduced motor nerve velocity, hypermyelination of retinal nerve fibres,
CC atrophy of upper cerebellar vermis, absence of Purkinje cells and
CC abnormal neuronal lipid storage. (I) can also be used to produce
CC antisense nucleic acids, is useful as molecular weight or chromosome
CC markers, to identify genetic disorders, as hybridisation probes or
CC primers, as an antigen, identify and express recombinant protein for
CC analysis, characterisation or therapeutic use, or as markers for tissues
CC in which the corresponding protein is expressed. Diagnostic methods from
CC the present invention can be used to identify subjects having or at risk
CC of developing a disease or disorder associated with aberrant expression
CC or activity of (I). The assays can be utilised to identify a subject
CC having or at risk of developing a disorder associated with spastin
CC protein or spastin gene expression or activity. The present sequence
CC encodes a mutated human spastin.
CC N.B. The present sequence is not given in the present specification but
CC is derived from the human spastin nucleotide sequence (AAH20174) as
CC stated on page 14.
XX
SQ Sequence 12793 BP; 4163 A; 2257 C; 2487 G; 3886 T; 0 other;

Query Match 100.0%; Score 500; DB 22; Length 12793;
Best Local Similarity 100.0%; Pred. No. 4.7e-239;
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 atgattacaggaagacatgtaactcagctgcagcttctaatacagaagatttcacg 60

|||||
Db 1 atgattacaggaagacatgtaactcagctgcagcttctaatacagaagatttcacg 60
Oy 61 tctatacagaagaaatgaatacatcttgcctgcgcagagaattgtttcaatgta 120
|||||
Db 61 tctatacagaagaaatgaatacatcttgcctgcgcagagaattgtttcaatgta 120
Oy 121 tccattgataaagaagaatcacccatctgtttcatgtcctaagatggtttgaaaa 180
|||||
Db 121 tccattgataaagaagaatcacccatctgtttcatgtcctaagatggtttgaaaa 180
Oy 181 tcttataatacttttccagagattgacttttttttttttttttttttttttttttt 240
|||||
Db 181 tcttataatacttttccagagattgacttttttttttttttttttttttttttttt 240
Oy 241 aactatacagaagagtcagacatgttgaactcatctagactcaggaattccatcgt 300
|||||
Db 241 aactatacagaagagtcagacatgttgaactcatctagactcaggaattccatcgt 300
Oy 301 agtcaatttagaagatgaatctgaagcacagcttccagaatttttagacagacattgaca 360
|||||
Db 301 agtcaatttagaagatgaatctgaagcacagcttccagaatttttagacagacattgaca 360
Oy 361 aaacttgagaggtttgtccttaaaaattagatgacatctatacaaacctcgctattaa 420
|||||
Db 361 aaacttgagaggtttgtccttaaaaattagatgacatctatacaaacctcgctattaa 420
Oy 421 aaatataatcatcaccatcaccaagtcgcttttcagagataatggagaagatgacatt 480
|||||
Db 421 aaatataatcatcaccatcaccaagtcgcttttcagagataatggagaagatgacatt 480
Oy 481 gcagaaatgtgtatcaaa 500
|||||
Db 481 gcagaaatgtgtatcaaa 500

RESULT 6
AAH20175
ID AAH20175 standard; DNA; 11493 BP.
XX
AC AAH20175;
XX
DT 09-AUG-2001 (first entry)
XX
DE Mouse spastin nucleotide sequence SEQ ID NO:3.
XX
KW Human; mouse; spastin; ARSACS; chromosome 13q11; identification;
KW autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;
KW neurodegenerative disease; reduced sensory nerve conduction; diagnosis;
KW reduced motor nerve velocity; hypermyelination of retinal nerve fibres;
KW atrophy of upper cerebellar vermis; absence of Purkinje cell;
KW abnormal neuronal lipid storage; genetic disorder; characterisation; ds.
XX
OS Mus musculus.
XX
FH Key Location/Qualifiers
FT CDS 1..11493
FT /*tag= a
FT /product= "spastin"
XX
PN WO200129266-A2.
XX
PD 26-APR-2001.
XX
PF 20-OCT-2000; 2000MO-US29130.
XX
PR 20-OCT-1999; 99US-0160588.
XX
PA (UYMC-) UNIV MCGILL.
XX (HOP1-) HOPITAL SAINTE-JUSTINE.
PI Hudson TJ, Engert J, Richter A;
XX

DR WPI; 2001-308494/32.
 DR P-PSDB; AAB97820.
 XX
 PT New isolated polynucleotide, encoding spastin gene, and polypeptides.
 PT useful for diagnosing autosomal recessive spastic ataxia of
 PT Charlevoix-Saguenay disease by detecting two point mutations in spastin
 PT gene sequence -
 PS
 PS Claim 1; Fig 8; 76pp; English.
 CC The present invention describes human and mouse spastin, and mutated
 CC human spastin (autosomal recessive spastic ataxia of Charlevoix-Saguenay
 CC (ARSA) gene sequences (I). The spastin gene has been mapped to
 CC chromosome 13q11. (I) have neuroprotective activities and can be used in
 CC gene therapy and as a spastin polypeptide agonists. (I), their fragments
 CC or their complements can be useful for assaying the presence of a nucleic
 CC acid molecule in a sample. (I) is useful for diagnosing or aiding in the
 CC diagnosis of an early onset neurodegenerative disease in an individual.
 CC The neurodegenerative disease comprises reduced sensory nerve conduction,
 CC reduced motor nerve velocity, hypermyelination of retinal nerve fibres,
 CC atrophy of upper cerebellar vermis, absence of Purkinje cells and
 CC abnormal neuronal lipid storage. (I) can also be used to produce
 CC antisense nucleic acids, is useful as molecular weight or chromosome
 CC markers, to identify genetic disorders, as hybridisation probes or
 CC primers, as an antigen, identify and express recombinant protein for
 CC analysis, characterisation or therapeutic use, or as markers for tissues
 CC in which the corresponding protein is expressed. Diagnostic methods from
 CC the present invention can be used to identify subjects having or at risk
 CC of developing a disease or disorder associated with aberrant expression
 CC or activity of (I). The assays can be utilised to identify a subject
 CC having or at risk of developing a disorder associated with spastin
 CC protein or spastin gene expression or activity. The present sequence
 CC encodes mouse spastin as given in the present invention.
 XX
 SQ Sequence 11493 BP; 3599 A; 2281 C; 2387 G; 3226 T; 0 other;
 Query Match 7.0%; Score 35; DB 22; Length 11493;
 Best Local Similarity 100.0%; Pred. No. 1.9e-07;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 203 gattgaattattatgatgagatgcacattatccc 237
 Db 127 gattgaattattatgatgagatgcacattatccc 161
 RESULT 7
 AAH20130
 ID AAH20130 standard; DNA: 23 BP.
 XX
 AC AAH20130;
 XX
 DT 09-AUG-2001 (first entry)
 XX
 DE Human spastin ORF PCR primer SEQ ID NO:23.
 XX
 KW Human; mouse; spastin; ARSACS; chromosome 13q11; identification;
 KW autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;
 KW neurodegenerative disease; reduced sensory nerve conduction; diagnosis;
 KW reduced motor nerve velocity; hypermyelination of retinal nerve fibre;
 KW atrophy of upper cerebellar vermis; absence of Purkinje cell;
 KW abnormal neuronal lipid storage; genetic disorder; characterisation;
 KW PCR primer; ss.
 XX
 OS Homo sapiens.
 OS Synthetic.
 XX
 PN WO200129266-A2.
 XX
 PD 26-APR-2001.
 XX
 PF 20-OCT-2000; 2000WO-US29130.
 XX

PR 20-OCT-1999; 9905-0160588.
 XX
 PA (UYMC-) UNIV MCGILL.
 PA (HOP1-) HOPITAL SAINTE-JUSTINE.
 XX
 PI Hudson TJ, Engert J, Richter A;
 XX
 DR WPI; 2001-308494/32.
 XX
 PT New isolated polynucleotide, encoding spastin gene, and polypeptides,
 PT useful for diagnosing autosomal recessive spastic ataxia of
 PT Charlevoix-Saguenay disease by detecting two point mutations in spastin
 PT gene sequence -
 PS
 PS Claim 23; Fig 7; 76pp; English.
 CC The present invention describes human and mouse spastin, and mutated
 CC human spastin (autosomal recessive spastic ataxia of Charlevoix-Saguenay
 CC (ARSA) gene sequences (I). The spastin gene has been mapped to
 CC chromosome 13q11. (I) have neuroprotective activities and can be used in
 CC gene therapy and as a spastin polypeptide agonists. (I), their fragments
 CC or their complements can be useful for assaying the presence of a nucleic
 CC acid molecule in a sample. (I) is useful for diagnosing or aiding in the
 CC diagnosis of an early onset neurodegenerative disease in an individual.
 CC The neurodegenerative disease comprises reduced sensory nerve conduction,
 CC reduced motor nerve velocity, hypermyelination of retinal nerve fibres,
 CC atrophy of upper cerebellar vermis, absence of Purkinje cells and
 CC abnormal neuronal lipid storage. (I) can also be used to produce
 CC antisense nucleic acids, is useful as molecular weight or chromosome
 CC markers, to identify genetic disorders, as hybridisation probes or
 CC primers, as an antigen, identify and express recombinant protein for
 CC analysis, characterisation or therapeutic use, or as markers for tissues
 CC in which the corresponding protein is expressed. Diagnostic methods from
 CC the present invention can be used to identify subjects having or at risk
 CC of developing a disease or disorder associated with aberrant expression
 CC or activity of (I). The assays can be utilised to identify a subject
 CC having or at risk of developing a disorder associated with spastin
 CC protein or spastin gene expression or activity. AAH20122 to AAH20173
 CC represent specifically claimed primers which can be used in diagnostic
 CC methods from the present invention.
 XX
 SQ Sequence 23 BP; 7 A; 7 C; 3 G; 6 T; 0 other;
 Query Match 4.6%; Score 23; DB 22; Length 23;
 Best Local Similarity 100.0%; Pred. No. 0.17;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 392 gatgacatatacaacatccgct 414
 Db 1 gatgacatatacaacatccgct 23
 RESULT 8
 AAT35105
 ID AAT35105 standard; cDNA: 721 BP.
 XX
 AC AAT35105;
 XX
 DT 21-NOV-1996 (first entry)
 XX
 DE Down-regulated senescence clone, SEND31 proposed full length sequence.
 XX
 KW Senescence related gene; expression; inhibition; acceleration; tomato;
 KW lettuce; cabbage; banana; strawberry; wheat; maize; rice; rape; soybean;
 KW delayed senescence; yield; protein content; quality; tolerance;
 KW increased senescence; desiccant; ss.
 XX
 OS Lycopersicon esculentum.
 XX
 PN WO9507993-A1.
 XX
 PD 23-MAR-1995.
 XX

XX 13-SEP-1994; 94MO-GB01990.
 PF 13-SEP-1993; 93GB-0018927.
 PR (ZENE) ZENECA LTD.
 XX
 XX Drake CR, Farrell A, Grierson D, Hosken SE, John I;
 PI Schuch WM, Smart CM, Thomas H;
 DR WPI: 1995-111361/17.
 XX
 PT DNA constructs which modify expression of senescence-related genes -
 PT useful to accelerate or inhibit senescence in crop plants
 XX
 PS Claim 10, Page 51; 78pp; English.
 XX
 CC The sequences given in AAT35095-133 are senescence related genes and
 CC fragments which were used in the construct of the invention which
 CC modifies the expression of at least one senescence related gene.
 CC Using these constructs senescence may be inhibited or accelerated
 CC in plants including tomato, lettuce, cabbage, banana, strawberry,
 CC wheat, maize, rice, rape or soybean. Delayed senescence may
 CC indirectly prolong the life of the plant, increase yield, increase
 CC protein content of fruits, improve quality of leafy vegetables.
 CC Improve tolerance to disease, drought or other stress. Increased
 CC senescence may more rapidly break down unwanted plant material and
 CC so avoid the use of desiccants on crops. This sequence represents the
 CC proposed full length sequence of the senescence down-regulating clone,
 CC PSEN31, which is also known as 1M4. PSEN31 is a cDNA of approx. 0.9 kb
 CC encoding a mRNA of approx. 1.0 kb. The mRNA encoded by PSEN31 is
 CC expressed in green leaves of tomato plants, but at the onset of
 CC senescence its expression is switched off. The PSEN31 sequence
 CC exhibits 100% homology to the tomato cDNA clone TAS14 which is inducible
 CC by salt stress and ABA in tomato seedlings. Southern analysis suggests
 CC that there is one gene per haploid genome. This gene is specifically
 CC reduced during tomato leaf senescence. PSEN31 is deposited as NCIMB
 CC 40576.
 XX
 SQ Sequence 721 BP; 222 A; 115 C; 192 G; 192 T; 0 other;

Query Match 3.8%; Score 19; DB 16; Length 721;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 459 agataatgagagaatgcc 477
 ||||||||||||||||
 Db 303 agataatgagagaatgcc 321

RESULT 9
 AAS83269
 ID AAS83269 standard; cDNA; 1543 BP.
 XX
 AC AAS83269;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE DNA encoding novel human diagnostic protein #19073.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX
 OS Homo sapiens.
 XX
 PN MO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001MO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI: 2001-639362/73.
 DR P-PSDB; ABG19082.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics; forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX
 PS Claim 1; SEQ ID No 19073; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 1543 BP; 489 A; 296 C; 374 G; 384 T; 0 other;

Query Match 3.8%; Score 19; DB 23; Length 1543;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 192 attttcagagattgac 210
 ||||||||||||||||
 Db 1389 attttcagagattgac 1407

RESULT 10
 AAH72857
 ID AAH72857 standard; cDNA; 1588 BP.
 XX
 AC AAH72857;
 XX
 DT 19-SEP-2001 (first entry)
 XX
 DE Human cervical cancer marker nucleic acid 4131.
 XX
 KW Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.
 KW Homo sapiens.
 XX
 OS Homo sapiens.
 XX
 PN MO200142467-A2.
 XX
 PD 14-JUN-2001.
 XX
 PF 08-DEC-2000; 2000MO-US33312.
 XX
 PF 08-DEC-1999; 99US-0169681.
 PR 21-DEC-1999; 99US-0171350.
 PR 14-MAR-2000; 2000US-0189315.

PR 12-MAY-2000; 2000US-0203791.
 PR 09-JUN-2000; 2000US-0210600.
 PR 21-JUL-2000; 2000US-0220114.
 XX
 PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
 XX
 PI Schlegel R, Deeds J, Berger A, Zhao X;
 XX
 DR WPI; 2001-375006/39.
 XX
 PT New isolated nucleic acid for diagnosing and treating cervical cancer
 PT and for assessing and detecting compounds for treating the cancer -
 XX
 PS Claim 1; Page 892-893; 1051pp; English.
 CC The invention relates to novel genes (AAH68727-AAH73383) associated with
 CC cervical cancer with cytostatic activity. The nucleic acids and encoded
 CC polypeptides are useful: to assess if a patient is afflicted with
 CC cervical cancer or has a pre-malignant condition; to monitor the
 CC progression of cervical cancer or a premalignant condition in a patient;
 CC and to select and/or assess the efficacy of a compound or therapy for
 CC inhibiting cervical cancer in a patient. The nucleic acids may also be
 CC useful for gene therapy.
 XX
 SQ Sequence 1588 BP; 511 A; 272 C; 356 G; 440 T; 9 other;
 XX
 Query Match 3.8%; Score 19; DB 22; Length 1588;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 192 attttcagagattgac 210
 |||||
 Db 1365 attttcagagattgac 1383
 XX
 RESULT 11
 AA74557/C
 ID AAV74557 standard; DNA; 3365 BP.
 XX
 AC AAV74557;
 XX
 DT 16-MAR-1999 (first entry)
 XX
 DE Staphylococcus aureus contig SEQ ID #246.
 XX
 KW Computer readable medium; vaccine; S.aureus infection; immunodetection;
 KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
 KW skin infection; surgical wound infection; scalded skin syndrome;
 KW toxic shock syndrome; ds.
 XX
 OS Staphylococcus aureus.
 XX
 FH Key Location/Qualifiers
 FT m1sc-feature 421..480
 FT /tag- a
 FT /note- "these bases represent a line of missing text in
 FT the sequence listing in the specification. They
 FT are included to maintain the nucleotide numbering
 FT given in the specification for this DNA sequence"
 FT m1sc-feature 2221..2280
 FT /tag- b
 FT /note- "these bases represent a line of missing text in
 FT the sequence listing in the specification. They
 FT are included to maintain the nucleotide numbering
 FT given in the specification for this DNA sequence"
 XX
 PN EP786519-A2.
 XX
 PD 30-JUL-1997.
 XX
 PF 07-JAN-1997; 97EP-0100117.
 XX

PR 05-JAN-1996; 96US-0009861.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA;
 PI Rosen CA;
 XX
 DR WPI; 1997-374922/35.
 XX
 PT Polynucleotide(s) and proteins derived from Staphylococcus aureus -
 PT stored on computer readable medium and used in the production of
 PT anti-S.aureus vaccines
 XX
 PS Claim 1; Page 1037-1039; 3271pp; English.
 CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences
 CC of the invention. The DNA sequences are recorded on a computer readable
 CC medium, preferably selected from a floppy or hard disk, random access
 CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
 CC the S.aureus DNA sequences allows putative functions to be assigned so
 CC that protein-encoding or regulatory regions of commercial, therapeutic or
 CC industrial importance can be obtained. Specifically, sequences which are
 CC likely to encode antigens have been identified and these polypeptides can
 CC be used in a vaccine composition against S.aureus infection. The
 CC polypeptides can also be used in a kit for the immunodetection of
 CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,
 CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
 CC skin and surgical wound infections, scalded skin syndrome, toxic shock
 CC syndrome, etc. Organisms transformed with the DNA sequences can be used
 CC for recombinant production of the polypeptides. The new DNA sequences
 CC (and their fragments) are useful as primers or probes for isolating
 CC homologues of any of the S.aureus DNA sequences contained on the
 CC computer readable medium.
 XX
 SQ Sequence 3365 BP; 1165 A; 536 C; 505 G; 1020 T; 139 other;
 XX
 Query Match 3.8%; Score 19; DB 18; Length 3365;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 415 tatttaaaatataatcat 433
 |||||
 Db 852 TATTAAAAAATATATTCAT 834
 XX
 RESULT 12
 ABL14512
 ID ABL14512 standard; cDNA; 4057 BP.
 XX
 AC ABL14512;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 38018.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ss.
 KW Drosophila melanogaster.
 OS
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PF Venter JC, Adams M, Li PWD, Myers EM;
 XX
 PI

XX WPI: 2001-656860/75.
 DR P-SDB; ABB70409.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Claim 1; SEQ ID NO 38018; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB16176-AB150511), expressed DNA
 CC sequences (AB157737-AB1572072) and the encoded proteins
 CC (AB157737-AB1572072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pcc_sequences.
 CC
 XX Sequence 4057 BP; 1264 A; 799 C; 851 G; 1143 T; 0 other;
 SQ

Query Match 3.8%; Score 19; DB 23; Length 4057;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 183 ttatatcatcttcaga 201
 ||||||||||||||||
 DB 963 ttatatcatcttcaga 1001

RESULT 13
 ABL34145/c
 ID ABL34145 standard; DNA: 7076 BP.
 XX
 AC ABL34145;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Human immune system associated gene SEQ ID NO: 2118.
 XX
 KW Human; immune system disease; cytosine methylation; antiasthmatic;
 KW antiarteriosclerotic; antianaemic; cytosatic; noctropic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW antineumatic; antiarthritic; antidiabetic; antipsoriatic;
 KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
 KW gene; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200200928-A2.
 XX
 PD 03-JAN-2002.
 XX
 PF 02-JUL-2001; 2001WO-EP07537.
 XX
 PR 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 XX
 PA (EPIC-) EPIGENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K;
 XX
 DR WPI: 2002-130909/17.
 XX
 PT Nucleic acid comprising fragment of chemically modified gene, useful
 PT for diagnosis and treatment of diseases associated with abnormal
 PT cytosine methylation -

XX
 PS Claim 1; SEQ ID NO 2118; 32pp + Sequence Listing; German.
 XX
 CC The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention.
 CC
 XX Sequence 7076 BP; 1994 A; 65 C; 1467 G; 3550 T; 0 other;
 SQ

Query Match 3.8%; Score 19; DB 24; Length 7076;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 179 aatcttatcatcttt 197
 ||||||||||||||||
 DB 4408 AATCTTATATACATTTT 4390

RESULT 14
 AAH12392
 ID AAH12392 standard; CDNA: 431 BP.
 XX
 AC AAH12392;
 XX
 DT 26-JUN-2001 (first entry)
 XX
 DE Human cDNA clone (3'-primer) SEQ ID NO:9227.
 XX
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
 XX
 OS Homo sapiens.
 XX
 PN EP1074617-A2.
 XX
 PD 07-FEB-2001.
 XX
 PF 28-JUL-2000; 2000EP-0116126.
 XX
 PR 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX
 DR WPI: 2001-318749/34.
 XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 XX
 PS Claim 3; SEQ ID 9227; 2537pp + CD ROM; English.
 XX
 CC The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification, where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a

CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences. AAB9246 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.

SQ Sequence 431 BP; 153 A; 84 C; 63 G; 126 T; 5 other;

Query Match 3.6%; Score 18; DB 22; Length 431;
 Best Local Similarity 100.0%; Pred. No. 53;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 413 cttattaaaaatatatt 430
 ||||||||||||||||
 Db 1 cttattaaaaatatatt 18

RESULT 15

AAF55109/C
 ID AAF55109 standard; DNA: 513 BP.

AC AAF55109;

DT 15-MAY-2001 (first entry)

DE Nucleotide sequence of a BASB119 polypeptide.

XX BASB119; immune response; bacterial infection; genetic immunization;

KW otitis media; pneumonia; sinusitis; nosocomial infection;

KM invasive disease; hearing loss; fluid accumulation; antibacterial; ss.

XX Moraxella catarrhalis.

OS Moraxella catarrhalis.

FT Key Location/Qualifiers
 CDS 1..513
 FT /*tag= a
 FT /product= "BASB119"
 FT /note= "no termination codon given"

XX WO200109336-A1.

XX 08-FEB-2001.

XX 31-JUL-2000; 2000MO-EP07363.

XX 03-AUG-1999; 99GB-0018302.

XX (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

XX Thonnard J;

XX WPI; 2001-159873/16.

XX P-PSDB; AAB67488.

XX New BASB119 polypeptides and polynucleotides from Moraxella catarrhalis

XX strain ATCC 43617, useful as therapeutic agents or vaccines against

XX bacterial infections, e.g. otitis media or pneumonia -

XX Claim 13; Page 65; 83pp; English.

XX The present sequence encodes a BASB119 polypeptide of Moraxella

XX catarrhalis strain ATCC43617. BASB119 polypeptides and polynucleotides

XX are useful for generating an immune response in an animal.. The

XX polypeptides may also be used as prophylactic agents of bacterial

CC infections, particularly M. catarrhalis infections in mammals,
 CC especially humans. The polynucleotides are useful in therapy or
 CC prophylaxis, particularly genetic immunization against these infections
 CC or diseases. These diseases include otitis media in infants or
 CC children, pneumonia in elders, sinusitis, nosocomial infections and
 CC invasive diseases, chronic otitis media with hearing loss, fluid
 CC accumulation in the middle ear, infection of the upper respiratory
 CC tract, or inflammation of the middle ear. The polypeptides or
 CC polynucleotides may also be employed as research reagents and materials
 CC for discovering treatments of and diagnostics for diseases,
 CC particularly human diseases. In particular, the polypeptides or
 CC polynucleotides are useful in the discovery and development of the
 CC antibacterial compounds, or for diagnosing diseases, staging of the
 CC disease, determining the response of an infectious organism to drugs.

SQ Sequence 513 BP; 173 A; 78 C; 111 G; 151 T; 0 other;

Query Match 3.6%; Score 18; DB 22; Length 513;
 Best Local Similarity 100.0%; Pred. No. 53;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 104 ttgattgtccaatggtat 121
 ||||||||||||||||
 Db 103 TTGATTGTCTCAATGCTAT 86

Search completed: May 22, 2002, 08:35:42
 Job time: 6562 sec



THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 22, 2002, 06:34:35 ; Search time 91.58 Seconds
(without alignments)
1341.087 Million cell updates/sec

Title: US-09-693-205-7_COPY_1_500

Sequence: 1 atgattaccaggaagacat.....gcagaattgtatcaaa 500

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 383533 seqs, 122816752 residues

Word size : 0

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

Issued_Patents_NA:*
1: /cgn2_6/prodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/prodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/prodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/prodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/prodata/1/ina/PTUS_COMB.seq:*
6: /cgn2_6/prodata/1/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	18	3.6	3736	4	US-09-600-776-10
2	17	3.4	305	4	US-09-020-956-66
3	17	3.4	305	4	US-09-030-607-66
4	17	3.4	305	4	US-09-439-313-66
5	17	3.4	2442	1	US-08-390-162-5
6	17	3.4	2442	1	US-08-685-945B-5
7	17	3.4	8700	2	US-08-392-625-16
8	17	3.4	8700	2	US-08-466-961A-16
9	17	3.4	8700	2	US-08-645-193B-18
10	17	3.4	8835	3	US-08-884-324-10
11	17	3.4	28994	3	US-08-884-324-14
12	16	3.2	57	6	US-09-025-580-19
13	16	3.2	57	6	5432082-6
14	16	3.2	449	4	US-08-946-914-35
15	16	3.2	677	4	US-08-998-416-1092
16	16	3.2	685	4	US-09-327-357-100
17	16	3.2	821	4	US-08-998-416-541
18	16	3.2	872	4	US-08-998-416-487
19	16	3.2	1341	1	US-08-180-209B-26
20	16	3.2	1341	1	US-08-385-745-26
21	16	3.2	1341	4	US-08-485-388-26
22	16	3.2	1341	4	US-08-474-853-26
23	16	3.2	1341	5	PCT-US94-02629-26
24	16	3.2	1505	4	US-09-130-616-178
25	16	3.2	1773	4	US-09-130-616-177
26	16	3.2	1936	3	US-08-946-914-7
27	16	3.2	2156	1	US-08-321-356-1

28	16	3.2	2156	1	US-08-321-356-3	Sequence 3, App1
29	16	3.2	2372	4	US-09-130-616-174	Sequence 174, App
30	16	3.2	3450	2	US-08-378-817A-9	Sequence 9, App1
31	16	3.2	3592	3	US-08-714-918-63	Sequence 63, App1
32	16	3.2	3592	4	US-09-265-315-63	Sequence 63, App1
33	16	3.2	3592	4	US-09-265-315-63	Sequence 63, App1
34	16	3.2	3592	4	US-09-265-315-63	Sequence 63, App1
35	16	3.2	3906	2	US-08-469-537A-102	Sequence 102, App
36	16	3.2	4060	1	US-08-164-292B-1	Sequence 1, App1
37	16	3.2	4060	1	US-08-164-292B-3	Sequence 3, App1
38	16	3.2	4060	1	US-08-164-292B-5	Sequence 5, App1
39	16	3.2	4060	1	US-08-164-292B-7	Sequence 7, App1
40	16	3.2	4060	3	US-08-845-623-1	Sequence 1, App1
41	16	3.2	4060	3	US-08-845-623-3	Sequence 3, App1
42	16	3.2	4060	3	US-08-845-623-5	Sequence 5, App1
43	16	3.2	4060	3	US-08-845-623-7	Sequence 7, App1
44	16	3.2	4060	3	US-08-815-927-1	Sequence 1, App1
45	16	3.2	4060	3	US-08-815-927-3	Sequence 3, App1

ALIGNMENTS

RESULT 1
US-09-600-776-10
; Sequence 10, Application US/09600776
; Patent No. 6326168
; GENERAL INFORMATION:
; APPLICANT: Yamanouchi Pharmaceutical Co., Ltd.
; TITLE OF INVENTION: A novel potassium channel protein
; FILE REFERENCE: Y9903-PCT
; CURRENT APPLICATION NUMBER: US/09/600,776
; CURRENT FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: JP P1998-011434
; PRIOR FILING DATE: 1998-01-23
; PRIOR APPLICATION NUMBER: JP P1998-346198
; PRIOR FILING DATE: 1998-12-04
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 10
; LENGTH: 3736
; TYPE: DNA
; ORGANISM: Rattus sp.
US-09-600-776-10

Query Match 3.6%; Score 18; DB 4; Length 3736;
Best Local Similarity 100.0%; Pred. No. 8.2;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 321 ctgaagcagcttcacg 338
|||
Db 3148 ctgaagcagcttcacg 3165

RESULT 2
US-09-020-956-66/c
; Sequence 66, Application US/09020956
; Patent No. 6261562
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillin, Davin C.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS
; NUMBER OF SEQUENCES: 178
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/020,956
FILING DATE: 09-FEB-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.427C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 66:
SEQUENCE CHARACTERISTICS:
LENGTH: 305 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
US-09-020-956-66

Query Match
3.4%: Score 17; DB 4; Length 305;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 413 ctatttaaaaaatat 429
|||||
DB 259 CTTATTAAAAAATATAT 243

RESULT 3
US-09-030-607-66/C
Sequence 66, Application US/09030607
Patent No. 6262245
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Dillon, Davin C.
TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS FO
NUMBER OF SEQUENCES: 224
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/030,607
FILING DATE: 25-FEB-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.427C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 66:
SEQUENCE CHARACTERISTICS:
LENGTH: 305 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
```

```
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
US-09-030-607-66

Query Match
3.4%: Score 17; DB 4; Length 305;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 413 ctatttaaaaaatat 429
|||||
DB 259 CTTATTAAAAAATATAT 243

RESULT 4
US-09-439-313-66/C
Sequence 66, Application US/09439313
Patent No. 6329505
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Dillon, Davin C.
APPLICANT: Mitcham, Jennifer L.
APPLICANT: Harlocker, Susan Louise
APPLICANT: Jiang Yuqiu
APPLICANT: Reed, Steven G.
APPLICANT: Kalos, Michael
APPLICANT: Fanger, Gary
APPLICANT: Retter, Mark
APPLICANT: Solk, John
APPLICANT: Day, Craig
TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
FILE REFERENCE: 210121.427C9
CURRENT APPLICATION NUMBER: US/09/439,313
CURRENT FILING DATE: 1999-11-12
NUMBER OF SEQ ID NOS: 575
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 66
LENGTH: 305
TYPE: DNA
ORGANISM: Homo sapien
US-09-439-313-66

Query Match
3.4%: Score 17; DB 4; Length 305;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 413 ctatttaaaaaatat 429
|||||
DB 259 CTTATTAAAAAATATAT 243

RESULT 5
US-08-390-162-5/C
Sequence 5, Application US/08390162
Patent No. 5576192
GENERAL INFORMATION:
APPLICANT: Ichikawa, Atsushi
APPLICANT: Narumiya, Shuh
TITLE OF INVENTION: Prostaglandin E Receptors, Their DNA and
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
ADDRESS: Dunner
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
```


MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/390,162
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/024179
FILING DATE: 23-FEB-1993
APPLICATION NUMBER: JP 036580-1992
FILING DATE: 24-FEB-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 064889-1992
FILING DATE: 23-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Fordis, Jean B.
REGISTRATION NUMBER: 32,984
REFERENCE/DOCKET NUMBER: 04221-0020-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-408-4000
TELEFAX: 202-408-4400
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2442 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-390-162-5

Query Match 3.4%; Score 17; DB 1; Length 2442;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Caps 0;

OY 185 tataatattttcaga 201
|||||
Db 1674 TATATACATTTTCAGA 1658

RESULT 6
US-08-685-945B-5/C
Sequence 5, Application US/08685945B
Patent No. 5804415
GENERAL INFORMATION:
APPLICANT: Ichikawa, Atsushi
APPLICANT: Natumaya, Shuh
TITLE OF INVENTION: prostoglandin E Receptors, Their DNA and
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
STREET: 1300 I Street, N.W., Suite 700
City: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/685,945B
FILING DATE: 22-JUL-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/024179
FILING DATE: 23-FEB-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 036580-1992

FILING DATE: 24-FEB-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 064889-1992
FILING DATE: 23-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Fordis, Jean B.
REGISTRATION NUMBER: 32,984
REFERENCE/DOCKET NUMBER: 04221-0020-02000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-408-4000
TELEFAX: 202-408-4400
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2442 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-685-945B-5

Query Match 3.4%; Score 17; DB 1; Length 2442;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Caps 0;

OY 185 tataatattttcaga 201
|||||
Db 1674 TATATACATTTTCAGA 1658

RESULT 7
US-08-392-625-16/C
Sequence 16, Application US/08392625
Patent No. 5837485
GENERAL INFORMATION:
APPLICANT: Entian, Karl-Dieter
APPLICANT: G tz, Friedrich
APPLICANT: Schnell, No. 5837485bert
APPLICANT: Augustin, Johannes
APPLICANT: Engelke, Germar
APPLICANT: Rosensteel, Ralf
APPLICANT: Kaletta, Cortina
APPLICANT: Klein, Cora
APPLICANT: Wieland, Bernd
APPLICANT: Kupke, Thomas
APPLICANT: Jung, G nther
APPLICANT: Kellner, Roland
TITLE OF INVENTION: Biosynthetic Process For The Preparation
OF Chemical Compounds
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox
STREET: 1100 New York Avenue
City: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/392,625
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/876,791
FILING DATE: 30-APR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Esmond, Robert W.
REGISTRATION NUMBER: 32,893
REFERENCE/DOCKET NUMBER: 0652.0980002

```
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 8700 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-392-625-16

Query Match
Best Local Similarity 100.0%; Pred. No. 26;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 415 tattaaaaatatattc 431
Db 574 TATTAAAAATATATTTC 558

RESULT 8
US-08-466-961A-16/c
Sequence 16, Application US/08466961A
Patent No. 5843709
GENERAL INFORMATION:
APPLICANT: Entlan, Karl-Dieter
APPLICANT: G t z, Friedrich
APPLICANT: Schnell, No. 5843709bert
APPLICANT: Augustin, Johannes
APPLICANT: Engelke, Gernar
APPLICANT: Rosensteln, Ralf
APPLICANT: Kaletta, Corlina
APPLICANT: Klein, Cora
APPLICANT: Wieland, Bernd
APPLICANT: Kupke, Thomas
APPLICANT: Jung, G nther
APPLICANT: Kellner, Roland
TITLE OF INVENTION: Biosynthetic Process for the Preparation of
TITLE OF INVENTION: Chemical Compounds
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESS: Sterne, Kessler, Goldstein & Fox
STREET: 1100 New York Avenue, NW
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/466,961A
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/392,625
FILING DATE: 22-FEB-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/876,791
FILING DATE: 30-APR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/784,234
FILING DATE: 31-OCT-1991
ATTORNEY/AGENT INFORMATION:
NAME: Esmond, Robert W.
REGISTRATION NUMBER: 32,893
REFERENCE/DOCKET NUMBER: 0652.0980004
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
```

```
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 8700 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-466-961A-16

Query Match
Best Local Similarity 100.0%; Pred. No. 26;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 415 tattaaaaatatattc 431
Db 574 TATTAAAAATATATTTC 558

RESULT 9
US-08-645-193B-18/c
Sequence 18, Application US/08645193B
Patent No. 5962253
GENERAL INFORMATION:
APPLICANT: Kupke, Thomas
APPLICANT: Gotz, Friedrich
APPLICANT: Kempler, Christoph
APPLICANT: Jung, Gunther
TITLE OF INVENTION: Oxidative Decarboxylation of Peptides
TITLE OF INVENTION: Catalyzed by Flavoprotein Epld
NUMBER OF SEQUENCES: 70
CORRESPONDENCE ADDRESS:
ADDRESS: Sterne, Kessler, Goldstein & Fox P.L.L.C.
STREET: 1100 New York Avenue, Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/645,193B
FILING DATE: 13-MAY-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Esmond, Robert W.
REGISTRATION NUMBER: 32,893
REFERENCE/DOCKET NUMBER: 0652.1540000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 8700 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: both
MOLECULE TYPE: CDNA
US-08-645-193B-18

Query Match
Best Local Similarity 100.0%; Pred. No. 26;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 415 tattaaaaatatattc 431
Db 574 TATTAAAAATATATTTC 558

RESULT 10
```

US-08-884-324-10
; Sequence 10, Application US/08884324
; Patent No. 6060283
; GENERAL INFORMATION:
; APPLICANT: Takano, OKURA
; APPLICANT: Kakui TORIGOE
; APPLICANT: Masashi KURIMOTO
; TITLE OF INVENTION: GENOMIC DNA ENCODING A POLYPEPTIDE CAPABLE
; TITLE OF INVENTION: OF INDUCING THE PRODUCTION OF INTERFERON-
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/884,324
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 185,305/96
; FILING DATE: 27-JUN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BROWDY, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: OKURA-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8835 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: human
; TISSUE TYPE: placenta
; FEATURE:
; NAME/KEY: Intron
; LOCATION: 1..8835
; IDENTIFICATION METHOD: E
; US-08-884-324-10

Query Match 3.44; Score 17; DB 3; Length 8835;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 86 ttctgacctgacagaga 102
|||||
Db 3604 TTCTGCCTGCGACAGAGA 3620

RESULT 11
US-08-884-324-14
; Sequence 14, Application US/08884324
; Patent No. 6060283
; GENERAL INFORMATION:
; APPLICANT: Takano, OKURA
; APPLICANT: Kakui TORIGOE
; APPLICANT: Masashi KURIMOTO
; TITLE OF INVENTION: GENOMIC DNA ENCODING A POLYPEPTIDE CAPABLE
; TITLE OF INVENTION: OF INDUCING THE PRODUCTION OF INTERFERON-
; NUMBER OF SEQUENCES: 35

CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/884,324
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 185,305/96
FILING DATE: 27-JUN-1996
ATTORNEY/AGENT INFORMATION:
NAME: BROWDY, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: OKURA-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 28994 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
ORIGINAL SOURCE:
ORGANISM: human
TISSUE TYPE: placenta
FEATURE:
NAME/KEY: 5'UTR
LOCATION: 1..15606
IDENTIFICATION METHOD: E
NAME/KEY: leader peptide
LOCATION: 15607..15685
IDENTIFICATION METHOD: S
NAME/KEY: Intron
LOCATION: 15686..17056
IDENTIFICATION METHOD: E
NAME/KEY: leader peptide
LOCATION: 17057..17068
IDENTIFICATION METHOD: S
NAME/KEY: Intron
LOCATION: 17069..20451
IDENTIFICATION METHOD: E
NAME/KEY: leader peptide
LOCATION: 20452..20468
IDENTIFICATION METHOD: S
NAME/KEY: mat peptide
LOCATION: 20469..20586
IDENTIFICATION METHOD: S
NAME/KEY: Intron
LOCATION: 20587..21920
IDENTIFICATION METHOD: E
NAME/KEY: mat peptide
LOCATION: 21921..22054
IDENTIFICATION METHOD: S
NAME/KEY: Intron
LOCATION: 22055..26827
IDENTIFICATION METHOD: E
NAME/KEY: mat peptide
LOCATION: 26828..27046
IDENTIFICATION METHOD: S
NAME/KEY: 3'UTR
LOCATION: 27047..28994
IDENTIFICATION METHOD: E

US-08-884-324-14

Query Match

3.4%; Score 17; DB 3; Length 28994;

Best Local Similarity 100.0%; Pred. No. 26;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 86 ttctgcctgcgcagaga 102

Db 10367 TTCTGCGCTGCGAGAGA 10383

RESULT 12

US-09-025-580-19/c

; Sequence 19, Application US/09025580

; Patent No. 6162613

; GENERAL INFORMATION:

; APPLICANT: Su, Michael Shln-San

; APPLICANT: Fox, Ted

; APPLICANT: Wilson, Keith Phillip

; APPLICANT: Germann, Ursula A.

; TITLE OF INVENTION: Methods For Designing Inhibitors of

; TITLE OF INVENTION: Serine/Threonine Kinases and Tyrosine Kinase

; NUMBER OF SEQUENCES: 37

; CORRESPONDENCE ADDRESSES:

; ADDRESSEE: Fish & Neave

; STREET: 1251 Avenue of the Americas

; CITY: New York

; STATE: NY

; COUNTRY: US

; ZIP: 10020

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/025,580

; FILING DATE:

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: Haley, James F.

; REGISTRATION NUMBER: 27,794

; REFERENCE/DOCKET NUMBER: VPI 97-104

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 596-9000

; TELEFAX: (212) 596-9090

; INFORMATION FOR SEQ ID NO: 19:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 37 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: other nucleic acid

; DESCRIPTION: /desc = "oligonucleotide"

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

US-09-025-580-19

Query Match

3.2%; Score 16; DB 4; Length 37;

Best Local Similarity 100.0%; Pred. No. 77;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 67 caaggaagtaatgaat 82

Db 35 CAAGGAAGTAATGAAT 20

RESULT 13

5432082-6/c

; Patent NO. 5432082

; APPLICANT: GALEOTTI, CESIRA, PALIA, EMANUELA, RAUGEI, GIOVANNI;

; BENSI, GIULIANO; MELLI, MARIA L.

; TITLE OF INVENTION: EXPRESSION AND SECRETION VECTOR IN YEASTS,

; USEFUL FOR PREPARING HETEROLOGOUS PROTEINS

; NUMBER OF SEQUENCES: 6

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/69,455

; FILING DATE: 01-JUN-1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 654,069; 69,329

; FILING DATE: 11-FEB-1991; 02-JUL-1987

; APPLICATION NUMBER: 69,329

; FILING DATE: 02-JUL-1987

; SEQ ID NO: 6;

; LENGTH: 57

5432082-6

Query Match

3.2%; Score 16; DB 6; Length 57;

Best Local Similarity 100.0%; Pred. No. 77;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 418 taaaaaatatcat 433

Db 16 TAAAAATATATTCAT 1

RESULT 14

US-08-946-914-35/c

; Sequence 35, Application US/08946914

; Patent No. 6027916

; GENERAL INFORMATION:

; APPLICANT: Ni, Jian

; APPLICANT: Gentz, Reiner L.

; APPLICANT: Ruben, Steven M.

; TITLE OF INVENTION: Galectin 8, 9, 10 and 10SV

; NUMBER OF SEQUENCES: 60

; CORRESPONDENCE ADDRESSES:

; ADDRESSEE: Sterne, Kessler, Goldstein, & Fox P.L.L.C.

; STREET: 1100 New York Ave., Suite 600

; CITY: Washington

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20005-3934

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/946,914

; FILING DATE: Herewith

; CLASSIFICATION: 530

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 60/028,093

; FILING DATE: 09-OCT-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: Steffe, Eric K.

; REGISTRATION NUMBER: 36,688

; REFERENCE/DOCKET NUMBER: 1488, 0560001/EKS/SGW

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 202-371-2600

; TELEFAX: 202-371-2540

; INFORMATION FOR SEQ ID NO: 35:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 449 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

US-08-946-914-35

Query Match

3.2%; Score 16; DB 3; Length 449;

Best Local Similarity 100.0%; Pred. No. 78;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 304 catttagacgatgaa 319
|||||
Db 54 CATTTAGACGATGAA 39

RESULT 15

US-08-998-416-1092/C
; Sequence 1092, Application US/08998416
; Patent No. 6239264
; GENERAL INFORMATION:
; APPLICANT: Philippsen, Peter
; APPLICANT: Pohlmann, Rainer
; APPLICANT: Steiner, Sabine
; APPLICANT: Mohr, Christine
; APPLICANT: Wendland, Jurgen
; APPLICANT: Knechtle, Philipp
; APPLICANT: Reblschung, Corinne
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSTYPI
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 1152
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6239264artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: No. 6239264tn Carolina
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/998,416
; FILING DATE: 24-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 0016/97
; FILING DATE: 31-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: PF/5-30306/A/GC1976
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 1092:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 677 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: PRG1666UP
; US-08-998-416-1092

Query Match 3.2%; Score 16; DB 4; Length 677;
Best Local Similarity 100.0%; Pred. No. 78;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 415 tatataaataatatt 430
|||||
Db 442 TATTAATAAATAATATT 427

Search completed: May 22, 2002, 08:29:03
Job time: 6868 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 22, 2002, 04:30:30 ; Search time 2968.03 Seconds
(without alignments)
2273.723 Million cell updates/sec

Title: US-09-693-205-7_COPY_1_500
Perfect score: 500
Sequence: 1 atgatttcaggaagaccat.....gcagaattcgttaacaa 500

Scoring table:
OLIGO_NUC
Gapop 60.0, Gapext 60.0

Searched: 13736207 segs, 6748477542 residues

Word size : 0

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST*
1: em_estba:*
2: em_esthum:*
3: em_estlin:*
4: em_estlinu:*
5: em_estlov:*
6: em_estlpl:*
7: em_estro:*
8: em_hlc:*
9: gb_estl:*
10: gb_estl2:*
11: gb_hlc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pin:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	176	35.2	241	9	BE008891
C 2	57	11.4	266	10	BF326199
C 3	42	8.4	746	12	BH025017
C 4	42	8.4	796	12	BH126217
C 5	35	7.0	558	12	AZ649875
C 6	22	4.4	202	9	BB561400
C 7	22	4.4	440	12	AZ066617
C 8	22	4.4	496	12	AZ644393
C 9	22	4.4	717	12	BH110525
C 10	20	4.0	170	10	BF397419
C 11	20	4.0	347	10	BF703038
C 12	20	4.0	630	10	BG599717
C 13	20	4.0	630	10	BG889484
C 14	20	4.0	686	9	AM018508
C 15	20	4.0	788	12	BH348265
C 16	20	4.0	755	10	BG593815
C 17	20	4.0	968	12	CNS06FVJ

18	19	3.8	189	9	BE152099
C 19	19	3.8	193	9	AA481759
C 20	19	3.8	220	9	AI275781
C 21	19	3.8	252	9	AL642337
C 22	19	3.8	271	10	R96076
C 23	19	3.8	280	9	AI823324
C 24	19	3.8	287	9	AI932736
C 25	19	3.8	292	9	AV295506
C 26	19	3.8	303	9	AV304343
C 27	19	3.8	333	10	H86539
C 28	19	3.8	328	10	BG941630
C 29	19	3.8	331	10	BG942635
C 30	19	3.8	342	9	AI468151
C 31	19	3.8	342	10	BG133892
C 32	19	3.8	351	9	AW262504
C 33	19	3.8	357	9	AI282243
C 34	19	3.8	358	9	AI042486
C 35	19	3.8	358	9	AI753035
C 36	19	3.8	360	9	AV194703
C 37	19	3.8	368	10	BG941160
C 38	19	3.8	372	10	N67665
C 39	19	3.8	375	10	N24075
C 40	19	3.8	377	9	AI025218
C 41	19	3.8	379	9	AI095227
C 42	19	3.8	379	10	W32532
C 43	19	3.8	381	9	AW204595
C 44	19	3.8	384	12	AQ129047
C 45	19	3.8	400	12	BH316467

ALIGNMENTS

RESULT 1
LOCUS BE008891 241 bp mRNA linear EST 05-JUN-2000
DEFINITION CM4-BN0161-040400-132-d08 BN0161 Homo sapiens cDNA, mRNA sequence.
VERSION BE008891.1 GI:8269124
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

JOURNAL
MEDLINE 20202663
COMMENT Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

TITLE
CONTACT: Simpson A.J.G.
LABORATORY OF Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2-CM4-BN0161-040400-132-d08&ts=2000-04-04&ta=1)
Seq primer: puc 18 forward
High quality sequence start: 6
High quality sequence stop: 241.
Location/Qualifiers
1. .241
/organism="Homo sapiens"

FEATURES
source

/db_xref="taxon:9606"
/clone_lib="BND0161"
/dev_stage="Adult"
/note="Organ: breast.normal; Vector: puc18; Site.1: SmaI;
Site.2: SmaI: A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

BASE COUNT 73 a 40 c 47 g 81 t
ORIGIN

Query Match 35.2%; Score 176; DB 9; Length 241;
Best Local Similarity 100.0%; Pred. No. 3.3e-81;
Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 275 ctcattagactcagatcttcattagcattagcagatgaatcgaagcagcgtt 334
|||||
Db 241 CTCATTAGACTCAGATCTTCATTAGCATTAGACATGATCGAAGCAGACACTT 182
|||||

QY 335 ccgaatttttagcagacatgttgcacaaactgtgaggtgttgccttaaaattagat 394
|||||
Db 181 CCGAAATTTTACGACATGTGACAAAACCTTGAGGGTTGCTTAAAAAATTAGAT 122
|||||

QY 395 gcatctatacaacatcgccttataaaatatattcattaccattaccagttgc 450
|||||
Db 121 GCATCTATACACATCCGCTTATTAAAAATATATTCATTACCATTCAGACTGC 66
|||||

RESULT 2
BF326199 266 bp mRNA linear EST 22-NOV-2000
LOCUS MRO-AN0083-160900-003-d10 AN0083 Homo sapiens cDNA, mRNA, sequence.
DEFINITION BF326199
ACCESSION BF326199
VERSION BF326199.1 GI:11296947
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 266)
Dias Neto, E., Garcia Correia, R., Verjovski-Almeida, S., Briones, M. R.,
Nagai, M. A., da Silva, M. Jr., Zago, M. A., Bordin, S., Costa, F. F.,
Goldman, G. H., Carvalho, A. F., Matsukuma, A., Baia, G. S., Simpson, D. H.,
Brunstein, A., de Oliveira, P. S., Bucher, P., Jongeneel, C. V., O'Hare,
M. J., Soares, F., Brentani, R. R., Reis, L. F., de Souza, S. J. and
Simpson, A. J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
CONTACT: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=MRO&cl=MRO-AN0083-
160900-003-d10&ts=2000-09-16&cl=1)
Seq primer: puc 18 forward
High quality sequence stop: 11.
Location/Qualifiers
1. 266
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="AN0083"
/dev_stage="Adult"

FEATURES

source

Query Match 8.4%; Score 42; DB 12; Length 746;

/note="Organ: amnion.normal; Vector: puc18; Site.1: SmaI;
Site.2: SmaI: A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

BASE COUNT 71 a 67 c 52 g 76 t
ORIGIN

Query Match 11.4%; Score 57; DB 10; Length 266;
Best Local Similarity 100.0%; Pred. No. 7.8e-19;
Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 aggaagacccttactagctgagctcttaattccagagcttgcagctcttt 66
|||||
Db 210 AGGAAGACCATGTACGCTGCACTTAAATCCGAGACGATTCACGCTTAT 266
|||||

RESULT 3
BH025017/c 746 bp DNA linear GSS 17-JUL-2001
LOCUS RPCI-24-318E15.TJ RPCI-24 Mus musculus genomic clone RPCI-24-318E15
DEFINITION
ACCESSION BH025017
VERSION BH025017.1 GI:14788481
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE
AUTHORS
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 746)
Zhao, S., Nieman, W., Malek, J., Shatsman, S., Akiret, B., Levins, M.,
Tsegaye, G., Geer, A., Krol, M., Shvartsbeyn, A., Georegeorgis, E.,
Russell, D., de Jong, P. and Fraser, C. M.
Mouse BAC End sequences from Library RPCI-24
Unpublished (1999)
Other GSSs: RPCI-24-318E15.TVB
CONTACT: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-24. For BAC
library availability, please contact Pieter de Jong
(pdejong@mail.cho.org). Clones may be purchased from BACPAC
Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end
page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 318 row: E column: 15
Seq primer: SP6
Class: BAC ends.
Location/Qualifiers
1. 746
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone=RPCI-24-318E15"
/clone_lib="RPCI-24"
/sex="Male"
/cell_type="Spleen/Brain"
/note="Vector: pPRBAC1; Site.1: BamHI; Site.2: BamHI;
RPCI-24 Mouse BAC library produced by Pieter de Jong. The
library was cloned in the pPRBAC1 cloning vector at the
BamHI sites using MboI partially digested male C57BL/6J
DNA."

FEATURES

source

BASE COUNT 229 a 142 c 137 g 238 t
ORIGIN

Query Match

Best Local Similarity 100.0%; Pred. No. 6.1e-11;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 7 tacaggaagacatgtactcagctgcagctcttaacccaga 48
|||||
Db 591 TACAGGAAGACATGTACTCAGCTGCAGCTTCTAAATCCAGA 550

RESULT 4
BH126217/c 796 bp DNA linear GSS 19-JUL-2001
LOCUS
DEFINITION
RPCI-24-318C15.TJ RPCI-24 Mus musculus genomic clone RPCI-24-318C15
, DNA sequence.
ACCESSION
VERSION
BH126217.1 GI:14969729
KEYWORDS
GSS.
SOURCE
house mouse.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 796)
Zhaio,S., Nieman,W., Malek,J., Shatsman,S., Akiret,B., Levins,M.,
Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
Russell,D., de Jong,P. and Fraser,C.M.
Mouse BAC End Sequences from Library RPCI-24
Unpublished (1999)
Other_GSSs: RPCI-24-318C15.TVB
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-24. For BAC
library availability, please contact Pieter de Jong
(pdejong@mail.cho.org). Clones may be purchased from BACPC
Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end
page: http://www.choi.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 318 row: C column: 15
Seq primer: SP6
Class: BAC ends.
Location/Qualifiers
1..796
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-318C15"
/clone_1lb="RPCI-24"
/sex="Male"
/cell_type="Spleen/Brain"
/note="Vector: pTRABAC1; Site_1: BamHI; Site_2: BamHI;
RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
library was cloned in the pTRABAC1 cloning vector at the
BamHI sites using MboI partially digested male C57BL/6J
DNA."
BASE COUNT 244 a 151 c 144 g 257 t
ORIGIN

Query Match 8.4%; Score 42; DB 12; Length 796;
Best Local Similarity 100.0%; Pred. No. 6.1e-11;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 7 tacaggaagacatgtactcagctgcagctcttaacccaga 48
|||||
Db 614 TACAGGAAGACATGTACTCAGCTGCAGCTTCTAAATCCAGA 573

RESULT 5
A2649875 558 bp DNA linear GSS 14-DEC-2000
LOCUS
DEFINITION 1M0519B14R Mouse 10kb plasmid UUGC1M library Mus musculus genomic

clone UUGC1M0519B14 R. DNA sequence.
ACCESSION
VERSION
A2649875.1 GI:11783794
KEYWORDS
GSS.
SOURCE
house mouse.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 558)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,
M., Rose,M., Rose,R., Stokes,R., Tingy,A., von Niederhausern,A.
and Wright,D., Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SIC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert length: 10000 Std Error: 0.00
Plate: 0519 row: B column: 14
Seq primer: CACACAGGAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 558.
Location/Qualifiers
1..558
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0519B14"
/clone_1lb="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lbb_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: pMD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(<http://www.jax.org/resources/documents/dnares/>). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adapter oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pMD42 (g114732114[9]AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
adapted vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
BASE COUNT 200 a 99 c 96 g 163 t
ORIGIN

Query Match 7.0%; Score 35; DB 12; Length 558;
Best Local Similarity 100.0%; Pred. No. 2.8e-07;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 203 gattgacttatttgatgagatgcacattatccc 237
|||||
Db 51 GATTGACTTATTGTGATGAGATGCACCTATGCC 85

RESULT 6
BB561400 202 bp mRNA linear EST 01-AUG-2000
LOCUS

DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS
B8561400 Riken full-length enriched, 10 days neonate olfactory brain Mus musculus CDNA clone E530213F07 3', mRNA sequence.	B8561400			B8561400.1 GI:9647766 EST.	house mouse. Mus musculus	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 302)	Kono, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carninci, P., Endo, T., Fukuda, S., Fukushima, Y., Hara, A., Hayatsu, N., Hirozane, T., Hoti, F., Ishii, T., Ishikawa, J., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N., Kiyosawa, H., Koijima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Ono, T., Owg, C., Saito, K., Sakai, C., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shinagawa, A., Shiraki, T., Sogabe, T., Sugihara, Y., Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomioka, N., Toya, T., Tsunoda, Y., Watabiki, A., Watanabe, S., Yamamura, T., Yamanaka, I., Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.
Riken Mouse ESTs (Kono, H., et al.).							
Unpublished (2000)							
Contact: Yoshihide Hayashizaki							

Contact: Yoshhide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center(GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-resgsc.riken.go.jp,
 URL: <http://genome.gsc.riken.go.jp/>
 Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasakawa,
 N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Thermostabilization and thermoactivation of the labile enzymes by
 trehalose and its application for the syntheses of full length
 cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
 Itoh, M., Kitsunai, T., Akiyama, J., Shibata, K., Iwawa, M., Kawai, J.,
 Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki,
 Y. and Hayashizaki, Y.
 Automated filtration-based high-throughput plasmid preparation
 system. Genome Res. 9 (5), 463-470 (1999)
 Carninci, P. and Hayashizaki, Y.
 High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
 19-44 (1999)
 Please visit our web site (<http://genome.riken.go.jp>) for
 further details.

FEATURES	Location/Qualifiers
source	1. .202

1. 102
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="E350213f07"
/clone_11b="RIKEN full-length enriched, 10 days neonate
olfactory brain"
/tissue_type="olfactory brain"
/dev_stage="10 days neonate"
/lab_host="pDH10p"
/note="Site_1: SalI. Site_2: BamHI, cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGGAGAGACGGCGCCCACTGACTTTTCTTTTCTTTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. Second strand cDNA was prepared with the
primer adapter of sequence [5'
GAGGAGAGAGATTCGACGTTAAATTAATTAATTAATCCCCCCCCCCC 3']. cDNA
was cleaved with BamHI and XhoI. Vector: a modified
bluescript KS(+) after bulk excision from Lambda FIC I."

BASE COUNT	62 a	53 c	29 g	57 t	1 others
ORIGIN					
Query Match	4.4%	Score 22:	DB 9:	Length 202:	
Best Local Similarity	100.0%	Pred. No. 1.7:			
Matches 22: Conservative	0:	Mismatches	0:	Indels	0:
Gaps	0:				
Oy	453	ttttgcagataatggagaagat	474		
Db	90	TTTTGCAGATTATGGAGAGAT	111		

LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS
7								
AZ066617/c								
	AZ066617	440 bp	DNA	linear	GSS 30-MAR-2000			
	RPCT-23-433L3.TV	RPCT-23	Mus musculus	genomic clone	RPCT-23-433L3,			
	DNA sequence.							
	AZ066617							
	AZ066617.1	GI:7357869						
	GSS.							
	house mouse.							
	Mus musculus							
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;							
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.							
	1 (bases 1 to 440)							
	Zhao, S., Niemman, W., Feldblum, T., Malek, J., Shatsman, S., Akintet							

TITLE	Mouse BAC End Sequences from Library RPCI-23
JOURNAL	Unpublished (1999)
COMMENT	Other_GSSs: RPCI-23-433L3.TJ

Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel.: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPci-23. For BAC library availability, please contact Pieter de Jong (pieterdejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.html>) or from Resea.ch Genetics (info@resgen.com). BAC end page: http://www.tigr.org/tdb/bac-ends/mouse/bac_end_intro.html

Plate: 433 row: L column: 3

Seq primer: T7

Class: BAC ends.

FEATURES	Location/Qualifiers
source	1. .440

```

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPC1-23-433L3"
/clone_1lb="RPC1-23"
/sex="Female"
/lab_host="DH10B"
/notes="Organ: Kidney/Brain; Vector: pBAC3.6; site_1:
EcoRI; site_2: EcoRI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcoRI and EcoRI Methylase. Size
selected DNA was cloned into the pBAC3.6 vector at the
EcoRI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies)."

```

Query Match	4.4%	Score 22;	DB 12;	Length 440;
Best Local Similarity	100.0%;	Pred. No. 1.8;		
Matches 22; Conservative	0;	Mismatches	0;	Gaps 0;
0y	453	ttttgcagataatgagagaagat	474	

```

Db      411 TTTTCAGATATGAGAGAT 390
|||||
RESULT  8
LOCUS   AZ644393
DEFINITION  496 bp DNA linear GSS 14-DEC-2000
clone UUGCIM0508M18 F, DNA sequence.
ACCESSION  AZ644393
VERSION    AZ644393.1 GI:11772878
KEYWORDS   house mouse.
SOURCE     Mus musculus
ORGANISM   Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 496)
AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausen,A.
and Wright,D., Weiss,R.
TITLE      Mouse whole genome scaffolding with paired end reads from 10kb
JOURNAL    Unpublished (2000)
COMMENT    Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SIC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunne@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0508 row: M column: 18
Seq primer: CGTTGTAACGACGCGCCACT
Class: plasmid ends
High quality sequence stop: 496.
Location/Qualifiers
1. 496
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGCIM0508M18"
/clone_11b="Mouse 10kb plasmid UUGCIM library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F.-"
/notes="Vector: pMD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pMD42 (g1147321149b1AF129072.1), a copy-number
inducible derivative of plasmid RL. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
adapted vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
BASE COUNT 181 a 94 c 85 g 136 t
ORIGIN
Query Match 4.4%; Score 22; DB 12; Length 496;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Oy      453 ttctgcagataatgagagaat 474
|||||
Db      162 TTTTCAGATATGAGAGAT 183
|||||
RESULT  9
LOCUS   BH110525/c
DEFINITION  717 bp DNA linear GSS:19-JUL-2001
RPCI-24-340G10-TV RPCI-24 Mus musculus genomic clone RPCI-24-340G10
, DNA sequence.
ACCESSION  BH110525
VERSION    BH110525.1 GI:14944731
KEYWORDS   GSS.
SOURCE     house mouse.
ORGANISM   Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 717)
AUTHORS   Zhao,S., Niernan,W., Malek,J., Shatsman,S., Akiret,B., Levins,M.,
Tregayle,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
Russell,D., de Jong,P. and Fraser,C.M.
Mouse BAC End Sequences from Library RPCI-24
Unpublished (1999)
Other GSSs: RPCI-24-340G10-TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-24. For BAC
library availability, please contact Pieter de Jong
(pdejong@mail.cho.org). Clones may be purchased from BACPAC
Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end
page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 340 row: G column: 10
Seq primer: SP6
Class: BAC ends.
Location/Qualifiers
1. 717
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-340G10"
/clone_11b="RPCI-24"
/sex="Male"
/cell_type="Spleen/Brain"
/notes="Vector: pPARBAC1; Site_1: BamHI; Site_2: BamHI.
RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
library was cloned in the pPARBAC1 cloning vector at the
BamHI sites using MboI partially digested male C57BL/6J
DNA."
BASE COUNT 206 a 125 c 133 g 253 t
ORIGIN
Query Match 4.4%; Score 22; DB 12; Length 717;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy      453 ttctgcagataatgagagaat 474
|||||
Db      539 TTTTCAGATATGAGAGAT 518
|||||
RESULT  10
LOCUS   BF397419/c
DEFINITION  170 bp mRNA linear EST 27-NOV-2000
UI-R-BS2-bel-b-03-0-UI.s1 UI-R-BS2 Rattus norvegicus cDNA clone
BF397419
ACCESSION  BF397419
VERSION    BF397419.1 GI:11382402

```

KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus

REFERENCE 1 (bases 1 to 170)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery

JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu

The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the normalized embryo at 13 dpc library cDNA Library Preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=yes.

FEATURES
source Location/Qualifiers
1..170
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UT-R-BS2-bel-b-03-0-UI"
/clone_lib="UT-R-BS2"
/dev_stage="embryonic 13 dpc"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker. Site.1: Not I; Site.2: Eco RI; The UT-R-BS2 library is a subtracted library derived from 13 dpc whole embryo tissue. For a detailed description of the library from which this clone was derived, please visit our web site at ratest.eng.uiowa.edu. The subtraction has been previously described in (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)
TAG_LIB="UT-R-BS2"
TAG_TISSUE="embryo at 13 dpc"
TAG_SEQ="ATATCC"

BASE COUNT 43 a 29 c 31 g 67 t
ORIGIN

Query Match 4.0%; Score 20; DB 10; Length 170;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 453 ttgtgcagataatgaggaag 472
|||||
Db 134 TTTTGCAGTATATGAGAG 115

RESULT 11
LOCUS BF703038 347 bp mRNA linear EST 22-DEC-2000
DEFINITION MI-P-E5-abn-e-02-1-UM.s1 MI-P-E5 Sus scrofa cDNA clone
VERSION BF703038
KEYWORDS BF703038.1 GI:11988446
SOURCE EST.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 347)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery

JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Tuggle CK
Molecular Genetics Laboratory, Department of Animal Science
Iowa State University
201 Kildee Hall, Ames, IA 50011-3150, USA
Tel: 5152944252
Fax: 5152942401
Email: cktuggle@iastate.edu

The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the non-normalized embryo at gestational day 12 library cDNA Library Preparation: RJ Woods, JA Green, RS Prather S142 Animal Science Research Center, Department of Animal Science, University of Missouri-Columbia, 65211 Clone distribution: clones will be available through Research Genetics (www.resgen.com) The following repetitive elements were found in this cDNA sequence: 1-38,
>AT_rich#low_complexity
Seq primer: M13 Forward
POLYA=yes.

FEATURES
source Location/Qualifiers
1..347
/organism="Sus scrofa"
/strain="crossbred"
/db_xref="taxon:9823"
/clone="MI-P-E5-abn-e-02-1-UM"
/clone_lib="MI-P-E5"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker. Site.1: Not I; Site.2: EcoRI; The MI-P-E5 library is derived from embryo at gestational day 12. For a detailed description of the library from which this clone was derived, please visit our web site at <http://pigest.genome.iastate.edu/>.
TAG_LIB="MI-P-E5"
TAG_TISSUE="embryo at gestational day 12"
TAG_SEQ="GTGAGA"

BASE COUNT 129 a 80 c 67 g 71 t
ORIGIN

Query Match 4.0%; Score 20; DB 10; Length 347;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 176 aaaaatcttatatacat 195
|||||
Db 12 AAAAATCTTATATACAT 31

RESULT 12
LOCUS BG599717 630 bp mRNA linear EST 12-APR-2001
DEFINITION EST504612 cSTS Solanum tuberosum cDNA clone cSTS26K9 5' sequence,
VERSION BG599717
KEYWORDS BG599717.1 GI:13616853
SOURCE EST.
ORGANISM Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
1 (bases 1 to 630)
van der Hoeven,R., Bezzerides,J., Sun,H., Cho,J., Chienlingo,A.,

double-stranded cDNA was ligated to Sal I adaptors (BRL), digested with Not I and cloned into the Not I and Sal I

sites of the pSPOR1 vector (BRU). Library was constructed by Matthew Clark (Lehrach lab; ICRF, London and Max Planck Institut fuer Molekulare Genetik, Berlin). cDNAs for EST analysis were selected following oligonucleotide hybridization fingerprinting of arrayed clones from zebrafish late somitogenesis (26 ss), adult liver or embryonic shield stage (5.6 h) libraries. Fingerprint data were used to computationally cluster cDNAs, and a single cDNA from each cluster was chosen for sequencing. In some cases multiple members of the same cluster were sequenced to assess clustering parameters or single clones were sequenced additional times to assess quality control."

BASE COUNT 200 a 148 c 173 g 162 t 3 others
ORIGIN

Query Match 4.0%; Score 20; DB 9; Length 686;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 310 agacgatgaatctgaagcac 329
|||||
Db 574 AGACGATGAATCTGAAGCAC 593

RESULT 15
BH348265/C 728 bp DNA linear GSS 03-DEC-2001
LOCUS CH230-120F14.TJ CHORI-230 Segment 1 Rattus norvegicus genomic clone
DEFINITION BH348265
ACCESSION BH348265
VERSION BH348265.1 GI:17278999
KEYWORDS GSS
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1 (bases 1 to 728)
AUTHORS Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K., Shvartsbeyn
A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de
Jong,P. and Fraser,C.M.
Rat BAC End Sequences from Library CHORI-230 EcoRI segment
Unpublished (1999)
other_GSSs: CH230-120F14.TJ
COMMENT
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org

Clones are derived from the rat BAC library CHORI-230
(http://www.chori.org/bacpac/rat230.htm). For BAC library
availability, please contact Pieter de Jong (pjejong@email.cho.org).
Clones may be purchased from BACPAC Resources
(http://www.chori.org/bacpac/or_ering_information.htm). BAC end
page: http://www.tigr.org/tdb/bac-ends/rat/bac_end_intro.html
Plate: 120 row: F column: 14
Seq primer: SP6
Class: BAC ends.

FEATURES
Source Location/Qualifiers

1..728
/organism="Rattus norvegicus"
/strain="BN/SSNHSd/MCW"
/db_xref="taxon:10116"
/clone="CH230-120F14"
/clone_lib="CHORI-230 Segment 1"
/sex="Female"
/cell_type="Brain"
/note="Vector: PTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI;
CHORI-230 Rat (BN/SSNHSd/MCW) BAC library produced by

Pieter de Jong"
BASE COUNT 273 a 103 c 145 g 207 t
ORIGIN

Query Match 4.0%; Score 20; DB 12; Length 728;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 415 tattaaataatattcatt 434
|||||
Db 115 TATTAAAAATATATTCATT 96

Search completed: May 22, 2002, 07:30:25
Job time: 10795 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 22, 2002, 08:35:42 ; Search time 373 Seconds
(without alignments)
6448.786 Million cell updates/sec

Title: US-09-693-205-7_COPY_5300_6700

Perfect score: 1401

Sequence: 1 gtacgcagtaaacactagagc.....aaagacacttaagaagt 1401

Scoring table:

OLIGO_NUC

Searched: 1736436 seqs, 858457221 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

N_Geneseq_032802:*
1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1401	100.0	12792	22	AAH20176 Human mutated spas
2	1295	92.4	12793	22	AAH20174 Human spastin nucl
3	1295	92.4	12793	22	AAH20178 Human mutated spas
4	1295	92.4	12793	22	AAH20179 Human mutated spas
5	1295	92.4	12793	22	AAH20182 Human mutated spas
6	53	3.8	11493	22	AAH20175 Mouse spastin nucl
7	24	1.7	24	22	AAH20151 Human spastin ORF
8	21	1.5	21	22	AAH20149 Human spastin ORF
9	21	1.5	21	22	AAH20150 Human spastin ORF

10	21	1.5	21	22	AAH20152
11	20	1.4	20	22	AAH20154
12	20	1.4	8496	22	AB125800 Drosophila melanog
13	19	1.4	389	22	AF66282 Novel human polynu
14	19	1.4	448	22	AA83961 DNA encoding novel
15	19	1.4	963	22	AA86530 C glutamicum codin
16	19	1.4	1288	22	AA83963 DNA encoding novel
17	19	1.4	1803	22	AA860658 Human polynucleoti
18	19	1.4	1960	22	AA158872 Human polynucleoti
19	19	1.4	2235	21	AA47414 Sequence encoding
20	19	1.4	2880	22	AAH73238 Human cervical can
21	19	1.4	9575	22	AA36912 Human cardiovascular
22	19	1.4	349980	22	AA68528 C glutamicum codin
23	18	1.3	394	22	AAH81581 Human differential
24	18	1.3	484	22	AA93377 Spinal cord tissue
25	18	1.3	973	22	AAH32263 Human olfactory re
26	18	1.3	1275	22	AAH52209 Human AFP protein
27	18	1.3	1425	21	AA16217 Human prostate can
28	18	1.3	1660	22	AA158832 Human polynucleoti
29	18	1.3	1750	22	AAH14806 Human CDNA sequenc
30	18	1.3	1785	19	AAV41902 Nucleotide sequenc
31	18	1.3	1806	19	AAV41901 Nucleotide sequenc
32	18	1.3	2017	22	AA160618 Human polynucleoti
33	18	1.3	2160	23	AA590268 DNA encoding novel
34	18	1.3	2333	21	AA47184 Arabidopsis thalia
35	18	1.3	3426	23	AB117098 Drosophila melanog
36	18	1.3	3514	23	AB102222 Drosophila melanog
37	18	1.3	3747	19	AAV05714 Maize retinoblasto
38	18	1.3	3747	19	AAV17081 Maize retinoblasto
39	18	1.3	5041	22	AAH18245 Drosophila melanog
40	18	1.3	6996	23	AB121622 Human immune syste
41	18	1.3	17869	24	AB132104 Genomic sequence #
42	18	1.3	21632	22	AA542019 Human musculoskele
43	18	1.3	32134	22	AA37170 Human ovarian and
44	18	1.3	32134	22	ABA07858 Human reproductive
45	18	1.3	32134	22	AA103669

ALIGNMENTS

RESULT 1	
ID	AAH20176 standard; DNA: 12792 BP.
AC	AAH20176:
XX	09-AUG-2001 (first entry)
DE	Human mutated spastin nucleotide sequence SEQ ID NO:7.
XX	
KW	Human; mouse; spastin; ARSACS; Chromosome 13q11; Identification;
KW	autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;
KW	neurodegenerative disease; reduced sensory nerve conduction; diagnosis;
KW	reduced motor nerve velocity; hypermyelination of retinal nerve fibre;
KW	atrophy of upper cerebellar vermis; absence of Purkinje cell;
KW	abnormal neuronal lipid storage; genetic disorder; Characterisation; ds.
OS	Homo sapiens.
OS	Synthetic.
PH	
FT	Key
FT	CDS
FT	Location/Qualifiers
FT	/*tag- a
FT	77..6604
FT	/product= "mutated spastin"
XX	
PN	W0200129266-A2.
PN	
PD	26-APR-2001.
XX	
PF	20-OCT-2000; 2000WO-US29130.
XX	
PR	20-OCT-1999; 99US-0160588.

Human spastin ORF
Human spastin ORF
Drosophila melanog
Novel human polynu
DNA encoding novel
C glutamicum codin
DNA encoding novel
Human polynucleoti
Human polynucleoti
Sequence encoding
Human cervical can
Human cardiovascular
C glutamicum codin
Human differential
Spinal cord tissue
Human olfactory re
Human AFP protein
Human prostate can
Human polynucleoti
Nucleotide sequenc
Nucleotide sequenc
Human polynucleoti
DNA encoding novel
Arabidopsis thalia
Drosophila melanog
Drosophila melanog
Maize retinoblasto
Maize retinoblasto
Drosophila melanog
Human immune syste
Genomic sequence #
Human musculoskele
Human ovarian and
Human reproductive

XX (UYMC-) UNIV MCGILL.
PA (HOP-) HOPITAL SAINTE-JUSTINE.
XX
XX
PI Hudson TJ, Engert J, Richter A;
XX WPI: 2001-308494/32.
DR P-PSDB: AAB97821.
XX
PT New isolated polynucleotide, encoding spastin gene, and polypeptides,
PT useful for diagnosing autosomal recessive spastic ataxia of
PT Charlevoix-Saguenay disease by detecting two point mutations in spastin
PT gene sequence -
XX
PS Claim 1: Page -: 76pp: English.
XX
XX The present invention describes human and mouse spastin, and mutated
CC human spastin (autosomal recessive spastic ataxia of Charlevoix-Saguenay
CC (ARSCS)) gene sequences (I). The spastin gene has been mapped to
CC chromosome 13q11. (I) have neuroprotective activities and can be used in
CC gene therapy and as a spastin polypeptide agonists. (I), their fragments
CC or their complements can be useful for assaying the presence of a nucleic
CC acid molecule in a sample. (I) is useful for diagnosing or aiding in the
CC diagnosis of an early onset neurodegenerative disease in an individual.
CC The neurodegenerative disease comprises reduced sensory nerve conduction,
CC reduced motor nerve velocity, hypermyelination of retinal nerve fibres,
CC atrophy of upper cerebellar vermis, absence of Purkinje cells and
CC abnormal neuronal lipid storage. (I) can also be used to produce
CC antisense nucleic acids, is useful as molecular weight or chromosome
CC markers, to identify genetic disorders, as hybridisation probes or
CC primers, as an antigen, identify and express recombinant protein for
CC analysis, characterisation or therapeutic use, or as markers for tissues
CC in which the corresponding protein is expressed. Diagnostic methods from
CC the present invention can be used to identify subjects having or at risk
CC of developing a disease or disorder associated with aberrant expression
CC or activity of (I). The assays can be utilised to identify a subject
CC having or at risk of developing a disorder associated with Spastin
CC protein or spastin gene expression or activity. The present sequence
CC encodes a mutated human spastin.
CC N.B. The present sequence is not given in the present specification but
CC is derived from the human spastin nucleotide sequence (AAB20174) as
CC stated on page 14.
XX
SQ Sequence 12792 BP: 4163 A; 2256 C; 2487 G; 3886 T; 0 other;

Query Match 100.0%; Score 1401; DB 22; Length 12792;
Best Local Similarity 100.0%; Fred. No. 0;
Matches 1401; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtgcagtaaaacttagagcagtcgccaagcgacacaaagccttagaagaatgcatcc 60
DB 5300 gtgcagtaaaacttagagcagtcgccaagcgacacaaagccttagaagaatgcatcc 5359

QY 61 aatgctgttttaacaacttggcacagaatttggcgagaagaataattgacacaga 120
DB 5360 aatgctgttttaacaacttggcacagaatttggcgagaagaataattgacacaga 5419

QY 121 attaagagatccttatgtatcatctcttgaagaagaatgttgaagccttccaa 180
DB 5420 attaagagatccttatgtatcatctcttgaagaagaatgttgaagccttccaa 5479

QY 181 aatgcgatgatgcaagcgacagaatctgttctgttggtttatcctaagcagatcca 240
DB 5480 aatgcgatgatgcaagcgacagaatctgttctgttggtttatcctaagcagatcca 5539

QY 241 gttgatagaataatttgatataagttggcccatctgcaagggcgacgacttgtgtac 300
DB 5540 gttgatagaataatttgatataagttggcccatctgcaagggcgacgacttgtgtac 5599

QY 301 aacaacacacatttacaagaatgatgtttagagaattcagaatcttggaaaagcgag 360
DB 5600 aacaacacacatttacaagaatgatgtttagagaattcagaatcttggaaaagcgag 5659

QY 361 aaagaggaataatccttaataaacctgacgatatggaatagatcaattctgttatcat 420
DB 5660 aaagaggaataatccttaataaacctgacgatatggaatagatcaattctgttatcat 5719

QY 421 atcacagactgcccattcttaattctgtgcaatgacatccgttgtattttgacatccat 480
DB 5720 atcacagactgcccattcttaattctgtgcaatgacatccgttgtattttgacatccat 5779

QY 481 gccagatatagcacacagggccacatccatagtcgccgaagcattgttaagatttgat 540
DB 5780 gccagatatagcacacagggccacatccatagtcgccgaagcattgttaagatttgat 5839

QY 541 gccagatttggacacagcttccagatgtcttcagatcttattcttgggaaccattttaa 600
DB 5840 gccagatttggacacagcttccagatgtcttcagatcttattcttgggaaccattttaa 5889

QY 601 ctgataattgcacaatgttcaagatttcctcttcgttaatgcaagaatgcaaaagtctcg 660
DB 5900 ctgataattgcacaatgttcaagatttcctcttcgttaatgcaagaatgcaaaagtctcg 5959

QY 661 gaatttcgtctgttccagcatcagaagaatggttccagaatctttggacaaactgcgc 720
DB 5960 gaatttcgtctgttccagcatcagaagaatggttccagaatctttggacaaactgcgc 6019

QY 721 tcagatgggcagaacttcaatgtttcttaatacacatgtaaaaaattctatttggaa 780
DB 6020 tcagatgggcagaacttcaatgtttcttaatacacatgtaaaaaattctatttggaa 6079

QY 781 atagataagagctactgtgagcttaaatgtcgtatcatcagtlaaaggccaatcacagat 840
DB 6080 atagataagagctactgtgagcttaaatgtcgtatcatcagtlaaaggccaatcacagat 6139

QY 841 gggagcagattggaagaagaacaatttcacatcgttgaattgatgttactcaaaag 900
DB 6140 gggagcagattggaagaagaacaatttcacatcgttgaattgatgttactcaaaag 6199

QY 901 aggcagctcaaaagacataccagttcaacaataaccatctctgtgactgaggaactc 960
DB 6200 aggcagctcaaaagacataccagttcaacaataaccatctctgtgactgaggaactc 6259

QY 961 gaaggaaatcttactacgtgtgctaatttgytaatagatcaggcctttcaaglatagagaa 1020
DB 6260 gaaggaaatcttactacgtgtgctaatttgytaatagatcaggcctttcaaglatagagaa 6319

QY 1021 gtatctaaagtgatcatatcagctcacaaagacagatattactctttccacggtgc 1080
DB 6320 gtatctaaagtgatcatatcagctcacaaagacagatattactctttccacggtgc 6379

QY 1081 ggaatagctgcctgcatattacccaactacaataaaacccatagggccttcgtttttg 1140
DB 6380 ggaatagctgcctgcatattacccaactacaataaaacccatagggccttcgtttttg 6439

QY 1141 cctcttcttcttggagactggtgcatttcacatgtaagtgagccacttgcactggaatcca 1200
DB 6440 cctcttcttcttggagactggtgcatttcacatgtaagtgagccacttgcactggaatcca 6499

QY 1201 gccagaaggaactgtgctgtaataatgaggttggtttggaaatgacttggaaatcaac 1260
DB 6500 gccagaaggaactgtgctgtaataatgaggttggtttggaaatgacttggaaatcaac 6559

QY 1261 agtttaatgacagatataatagctcctgcataatgtaagttgcaatlaacagtttaaaaaac 1320
DB 6560 agtttaatgacagatataatagctcctgcataatgtaagttgcaatlaacagtttaaaaaac 6619

QY 1321 gtaattccctgtgttcgtatccacaacatlatcagtggttacagaacacccctatcatgttg 1380
DB 6620 gtaattccctgtgttcgtatccacaacatlatcagtggttacagaacacccctatcatgttg 6679

QY 1381 taaaggacactttaaagaagt 1401
DB 6680 taaaggacactttaaagaagt 6700

RESULT 2
AAH20174
ID AAH20174 standard; DNA: 12793 BP.
XX
AC AAH20174;
XX
09-AUG-2001 (first entry)
XX
Human spastin nucleotide sequence SEQ ID NO:1.
XX
Human; mouse; spastin; ARSACS; chromosome 13q11; identification;
KM autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;
KM neurodegenerative disease; reduced sensory nerve conduction; diagnosis;
KM reduced motor nerve velocity; hypermyelination of retinal nerve fibre;
KM atrophy of upper cerebellar vermis; absence of Purkinje cell;
XX abnormal neuronal lipid storage; genetic disorder; characterisation; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 77..11566
FT /tag- a
FT /product- "spastin"
XX
PN MO200129266-A2.
XX
PD 26-APR-2001.
XX
PF 20-OCT-2000; 2000MO-US29130.
XX
PR 20-OCT-1999; 99US-0160588.
XX
PA (UYMC-) UNIV MCGILL.
PA (HOP1-) HOPITAL SAINTE-JUSTINE.
XX
PI Hudson TJ, Engert J, Richter A;
XX
DR WPI: 2001-308494/32.
DR P-PSDB; AAB97819.
XX
XX
PT New isolated polynucleotide, encoding spastin gene, and polypeptides,
PT useful for diagnosing autosomal recessive spastic ataxia of
PT Charlevoix-Saguenay disease by detecting two point mutations in spastin
PT gene sequence -
XX
XX
PS Claim 1; Fig 9; 76pp; English.
XX
XX The present invention describes human and mouse spastin, and mutated
CC human spastin (autosomal recessive spastic ataxia of Charlevoix-Saguenay
CC (ARSACS)) gene sequences (I). The spastin gene has been mapped to
CC chromosome 13q11. (I) have neuroprotective activities and can be used in
CC gene therapy and as a spastin polypeptide agonists. (I), their fragments
CC or their complements can be useful for assaying the presence of a nucleic
CC acid molecule in a sample. (I) is useful for diagnosing or aiding in the
CC diagnosis of an early onset neurodegenerative disease in an individual.
CC The neurodegenerative disease comprises reduced sensory nerve conduction,
CC reduced motor nerve velocity, hypermyelination of retinal nerve fibres,
CC atrophy of upper cerebellar vermis, absence of Purkinje cells and
CC abnormal neuronal lipid storage. (I) can also be used to produce
CC antisense nucleic acids, is useful as molecular weight or chromosome
CC markers, to identify genetic disorders, as hybridisation probes or
CC primers, as an antigen, identify and express recombinant protein for
CC analysis, characterisation or therapeutic use, or as markers for tissues
CC in which the corresponding protein is expressed. Diagnostic methods from
CC the present invention can be used to identify subjects having or at risk
CC of developing a disease or disorder associated with aberrant expression
CC or activity of (I). The assays can be utilised to identify a subject
CC having or at risk of developing a disorder associated with spastin
CC protein or spastin gene expression or activity. The present sequence
CC encodes human spastin as given in the present invention.
XX
XX Sequence 12793 BP; 4163 A; 2256 C; 2487 G; 3887 T; 0 other;

Query Match 92.4%; Score 1295; DB 22; Length 12793;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1295; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gtagcagtaaacactagagcagctcccaagcgacacaaagcccttagaagatgcatcc 60
|||||
Db 5300 gtagcagtaaacactagagcagctcccaagcgacacaaagcccttagaagatgcatcc 5359
QY 61 aatgtctgttttacacacacttgagcagagaatttggcgagaagaanaattgacagcaga 120
|||||
Db 5360 aatgtctgttttacacacacttgagcagagaatttggcgagaagaanaattgacagcaga 5419
QY 121 attaagagatcccttaatgcatatccctctgaaaaagaaatgttgaaagactttcaa 180
|||||
Db 5420 attaagagatcccttaatgcatatccctctgaaaaagaaatgttgaaagactttcaa 5479
QY 181 aatgtctgtgtgacaaagcgacagaaatctgtttgtgttgaatcccttagacagatcca 240
|||||
Db 5480 aatgtctgtgtgacaaagcgacagaaatctgtttgtgttgaatcccttagacagatcca 5539
QY 241 gtgtagaataatgtgataaagtggcccccatgtgcaaggcgacacttgtgtgtac 300
|||||
Db 5540 gtgtagaataatgtgataaagtggcccccatgtgcaaggcgacacttgtgtgtac 5599
QY 301 aacacaccagcatcttacagaagaatgagttagaggaattcagaatcttgaaagacag 360
|||||
Db 5600 aacacaccagcatcttacagaagaatgagttagaggaattcagaatcttgaaagacag 5659
QY 361 aaagagggaaatcccttataaaacttgacagatggaatggaatcattctgtgtatcat 420
|||||
Db 5660 aaagagggaaatcccttataaaacttgacagatggaatggaatcattctgtgtatcat 5719
QY 421 atcacagactgccatcttatttctgtgcaatgacatccctgtgattttgatacctcat 480
|||||
Db 5720 atcacagactgccatcttatttctgtgcaatgacatccctgtgattttgatacctcat 5779
QY 481 gccagatatgcacacaggggccacatcatagtcctccggacgcatgtttagagatttgat 540
|||||
Db 5780 gccagatatgcacacaggggccacatcatagtcctccggacgcatgtttagagatttgat 5839
QY 541 gcaagatttgaagaacagtttctgaatgttctgtgacttttctgtggaccattttaa 600
|||||
Db 5840 gcaagatttgaagaacagtttctgaatgttctgtgacttttctgtggaccattttaa 5899
QY 601 ctggaataatgcacaatctcagatttccctctgttaatgacgaatgtgcacaaatttcg 660
|||||
Db 5900 ctggaataatgcacaatctcagatttccctctgttaatgacgaatgtgcacaaatttcg 5959
QY 661 gaaatttcgtctgtccagcatcagacagaatgtgtccagaatcttttggacaaactgagc 720
|||||
Db 5960 gaaatttcgtctgtccagcatcagacagaatgtgtccagaatcttttggacaaactgagc 6019
QY 721 tcaagatggggcagaactcttaatgttttcttatacacaatggaaaaaatttctattgtgaa 780
|||||
Db 6020 tcaagatggggcagaactcttaatgttttcttatacacaatggaaaaaatttctattgtgaa 6079
QY 781 atagataagatcagcagcctcaaatgtgtgtatcttagtaaaaggcaaaatccacagat 840
|||||
Db 6080 atagataagatcagcagcctcaaatgtgtgtatcttagtaaaaggcaaaatccacagat 6139
QY 841 ggagacagattgaaaaggaaacaaatttcacatgtcatctgtaattgatagttactaaag 900
|||||
Db 6140 ggagacagattgaaaaggaaacaaatttcacatgtcatctgtaattgatagttactaaag 6199
QY 901 aggcagcctcaagcagcatccagttcaacaaataactatactatgagatcggagagcct 960
|||||
Db 6200 aggcagcctcaagcagcatccagttcaacaaataactatactatgagatcggagagcct 6259
QY 961 gaaggaatcttactacgtgtgctcaattgttataagatcaggcttttcaagatgagaa 1020
|||||
Db 6260 gaaggaatcttactacgtgtgctcaattgttataagatcaggcttttcaagatgagaa 6319


```
QY 721 tcagatgggcagagacttcaatgtttcttaatcacatggaataattctattgtgaa 780
    |||||||
Db 6020 tcagatgggcagagacttcaatgtttcttaatcacatggaataattctattgtgaa 6079
QY 781 atagataagagctatgagctcttaaatgctgtatctcagtaaaggycnaaatcacagat 840
    |||||||
Db 6080 atagataagagctatgagctcttaaatgctgtatctcagtaaaggycnaaatcacagat 6139
QY 841 gggagcagattgtaaaagaaacaattctacatgcatctgttaattgtatgttactcaaaaag 900
    |||||||
Db 6140 gggagcagattgtaaaagaaacaattctacatgcatctgttaattgtatgttactcaaaaag 6199
QY 901 aggcagctcaagaacatcacccagttcaacaataaccatactatggagctaggagactct 960
    |||||||
Db 6200 aggcagctcaagaacatcacccagttcaacaataaccatactatggagctaggagactct 6259
QY 961 gaaagaaatctactacgttgcttaattgttaataagatcagagctttccaagtatggagaaa 1020
    |||||||
Db 6260 gaaagaaatctactacgttgcttaattgttaataagatcagagctttccaagtatggagaaa 6319
QY 1021 gtatctaaaaggtcatatatacagctcaagaacacaaatatattactctttccacgtgtgt 1080
    |||||||
Db 6320 gtatctaaaaggtcatatatacagctcaagaacacaaatatattactctttccacgtgtgt 6379
QY 1081 gggatgactgctgcattactactacatactataaaacccataaggccttctgtttttttg 1140
    |||||||
Db 6380 gggatgactgctgcattactactacatactataaaacccataaggccttctgtttttttg 6439
QY 1141 cctcttctcttgagagactggtggtccatttcatgtgaatggccacttgcacgtgattca 1200
    |||||||
Db 6440 cctcttctcttgagagactggtggtccatttcatgtgaatggccacttgcacgtgattca 6499
QY 1201 gccagaaggaactgtgtgctgtatgataaatggagttgtgttcgaagtgcagtgaataac 1260
    |||||||
Db 6500 gccagaaggaactgtgtgctgtatgataaatggagttgtgttcgaagtgcagtgaataac 6559
QY 1261 agtttaatgacagcattaatagctcccgcatatgt 1295
    |||||||
Db 6560 agtttaatgacagcattaatagctcccgcatatgt 6594

RESULT 4
ID AAH20179 standard; DNA: 12793 BP.
XX
AC AAH20179;
XX
DT 09-AUG-2001 (first entry)
XX
DE Human mutated spastin nucleotide sequence SEQ ID NO:12.
XX
KW Human; mouse; spastin; ARSACS; chromosome 13q11; identification;
KW autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;
KW neurodegenerative disease; reduced sensory nerve conduction; diagnosis;
KW reduced motor nerve velocity; hypermyelination of retinal nerve fibres;
KW atrophy of upper cerebellar vermis; absence of Purkinje cell;
KW abnormal neuronal lipid storage; genetic disorder; characterisation; ds.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN MO200129266-A2.
XX
PD 26-APR-2001.
XX
PF 20-OCT-2000; 2000MO-US29130.
XX
PR 20-OCT-1999; 99US-0160588.
XX
PA (UYMC-) UNIV MCGILL.
PA (HOP1-) HOPITAL SAINTE-JUSTINE.
XX
XX Hudson TJ, Engert J, Richter A;
PI
```

```
XX
DR WPI: 2001-308494/332.
XX
PT New isolated polynucleotide, encoding spastin gene, and polypeptides,
PT useful for diagnosing autosomal recessive spastic ataxia of
PT Charlevoix-Saguenay disease by detecting two point mutations in spastin
PT gene sequence -
XX
PS Claim 1; Page -: 76pp; English.
XX
CC The present invention describes human and mouse spastin, and mutated
CC human spastin (autosomal recessive spastic ataxia of Charlevoix-Saguenay
CC (ARSACS)) gene sequences (I). The spastin gene has been mapped to
CC chromosome 13q11. (I) have neuroprotective activities and can be used in
CC gene therapy and as a spastin polypeptide agonists. (I), their fragments
CC or their complements can be useful for assaying the presence of a nucleic
CC acid molecule in a sample. (I) is useful for diagnosing or aiding in the
CC diagnosis of an early onset neurodegenerative disease in an individual.
CC The neurodegenerative disease comprises reduced sensory nerve conduction,
CC reduced motor nerve velocity, hypermyelination of retinal nerve fibres,
CC atrophy of upper cerebellar vermis, absence of Purkinje cells and
CC abnormal neuronal lipid storage. (I) can also be used to produce
CC antisense nucleic acids, is useful as molecular weight or chromosome
CC markers, to identify genetic disorders, as hybridisation probes or
CC primers, as an antigen, identify and express recombinant protein for
CC analysis, characterisation or therapeutic use, or as markers for tissues
CC in which the corresponding protein is expressed. Diagnostic methods from
CC the present invention can be used to identify subjects having or at risk
CC of developing a disease or disorder associated with aberrant expression
CC or activity of (I). The assays can be utilised to identify a subject
CC having or at risk of developing a disorder associated with Spastin
CC protein or spastin gene expression or activity. The present sequence
CC encodes a mutated human spastin.
CC N.B. The present sequence is not given in the present specification but
CC is derived from the human spastin nucleotide sequence (AAH20174) as
CC stated on page 14.
XX
SQ Sequence 12793 BP; 4163 A; 2257 C; 2487 G; 3886 T; 0 other;

Query Match 92.4%; Score 1295; DB 22; Length 12793;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1295; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtagcagtaaaactagagcagctcccaagcgacacaaagccttagaagatagcatcc 60
    |||||||
Db 5300 gtagcagtaaaactagagcagctcccaagcgacacaaagccttagaagatagcatcc 5359
QY 61 aatgtctgttttacaacactgtggcacaagaatttggcgagaagaaattgacccgcaga 120
    |||||||
Db 5360 aatgtctgttttacaacactgtggcacaagaatttggcgagaagaaattgacccgcaga 5419
QY 121 atgaagagcatccttaatgcatatcctctctgaaagaaagcttgaagaccttctcaa 180
    |||||||
Db 5420 atgaagagcatccttaatgcatatcctctctgaaagaaagcttgaagaccttctcaa 5479
QY 181 aatcgtatgatgcaaaagcgacagaaatcgtgtttgtgttgatcctagacagcatca 240
    |||||||
Db 5480 aatcgtatgatgcaaaagcgacagaaatcgtgtttgtgttgatcctagacagcatca 5539
QY 241 gttagataaatatttgatgaatgtggcccatgtgcaaggcgccagccttgtgtgtac 300
    |||||||
Db 5540 gttagataaatatttgatgaatgtggcccatgtgcaaggcgccagccttgtgtgtac 5599
QY 301 aacacaccgcatcttacagaagatgatltagaggaattcgaatcttgaagaaagcacg 360
    |||||||
Db 5600 aacacaccgcatcttacagaagatgatltagaggaattcgaatcttgaagaaagcacg 5659
QY 361 aaagagggaaatccttataaaactggacagcatggaatagatcattctgtatcat 420
    |||||||
Db 5660 aaagagggaaatccttataaaactggacagcatggaatagatcattctgtatcat 5719
QY 421 atcacagactgccatcttattcttgcaatgacatccctgtatatttgatcccat 480
```

```

|||||
Db 5720 atcacagactgcccacatttatttcgtgcaatgacacactcgtgtattttgtcccat 5779
Qy 481 gccagatgacacaggggacacatccatagaccggagcagctgtttaagatttgat 540
Db 5780 gccagatacgacacaggggacacatccatagaccggagcagctgtttaagatttgat 5839
Qy 541 gccagattttaggacacagcttctcagatgltctgcatcttctatctcgtggaacccatttaa 600
Db 5840 gccagattttaggacacagcttctcagatgltctgcatcttctatctcgtggaacccatttaa 5839
Qy 601 ctgagataattgcaacaatglttcagatttctcttcgttaatgacagaatgacaaaagtttcg 660
Db 5900 ctgagataattgcaacaatglttcagatttctcttcgttaatgacagaatgacaaaagtttcg 5959
Qy 661 gaatttcgtctgttcagatcagacagatgttcagagattttttggacaaactgcgc 720
Db 5960 gaatttcgtctgttcagatcagacagatgttcagagattttttggacaaactgcgc 6019
Qy 721 tcagatggggcagaactctctaattcttcttaatacacaatgaaataattctatttgcga 780
Db 6020 tcagatggggcagaactctctaattcttcttaatacacaatgaaataattctatttgcga 6079
Qy 781 atagataagagtaactcgtgagcttaaatgltgctgtattcagtaaaaggacaaatcacagat 840
Db 6080 atagataagagtaactcgtgagcttaaatgltgctgtattcagtaaaaggacaaatcacagat 6139
Qy 841 ggagagaagattgaaaaagaaacatttctcagctcgtgtaattatagtggttactaaag 900
Db 6140 ggagagaagattgaaaaagaaacatttctcagctcgtgtaattatagtggttactaaag 6199
Qy 901 aggcagctcaaaagacataccagctcaacaataactatactatgtaactgaggaactc 960
Db 6200 aggcagctcaaaagacataccagctcaacaataactatactatgtaactgaggaactc 6259
Qy 961 gaagagaatcttactcgtggttaatttgaatagatcaggcttctcaagtatggagaa 1020
Db 6260 gaagagaatcttactcgtggttaatttgaatagatcaggcttctcaagtatggagaa 6319
Qy 1021 gttctcaaaagtgctatagcttcaagaagaacagatattactctttccacagcgtc 1080
Db 6320 gttctcaaaagtgctatagcttcaagaagaacagatattactctttccacagcgtc 6379
Qy 1081 ggaatgagctgcgtcacttactcacaactataaaaccccatagggcctctgtttttg 1140
Db 6380 ggaatgagctgcgtcacttactcacaactataaaaccccatagggcctctgtttttg 6439
Qy 1141 cctcttcttcttgagactcgtggtcatttcatgtaatgacacatttgacatgattca 1200
Db 6440 cctcttcttcttgagactcgtggtcatttcatgtaatgacacatttgacatgattca 6499
Qy 1201 gccagagaagaacctgtggcgtgtaataatgaggtggtgttgcgaatgactggaataac 1260
Db 6500 gccagagaagaacctgtggcgtgtaataatgaggtggtgttgcgaatgactggaataac 6559
Qy 1261 agtttaatgacagcatataatagctcctgcataagt 1295
Db 6560 agtttaatgacagcatataatagctcctgcataagt 6594

```

RESULT 5
AAH20182
ID AAH20182 standard; DNA; 12793 BP.

XX AAH20182;

DT 09-AUG-2001 (first entry)

DE Human mutated spastin nucleotide sequence SEQ ID NO:15.

KM Human; mouse; spastin; ARSACS; chromosome 13q11; identification;
KW Autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;
neurodegenerative disease; reduced sensory nerve conduction; diagnosis;

KM reduced motor nerve velocity; hypermyelination of retinal nerve fibre;
KW atrophy of upper cerebellar vermis; absence of Purkinje cell;
KM abnormal neuronal lipid storage; genetic disorder; characterisation; ds.
OS Homo sapiens.
OS Synthetic.
FH Key Location/Qualifiers
FT CDS 77..11566
FT /tag= a
FT /product= "mutated spastin"
XX WO200129266-A2.
XX 26-APR-2001.
XX 20-OCT-2000; 2000MO-US29130.
XX 20-OCT-1999; 9905-0160588.
XX (UYMC-) UNIV MCGILL.
PA (HOPIT-) HOPITAL SAINTE-JUSTINE.
PI Hudson TJ, Engert J, Richter A;
XX WPI; 2001-308494/32.
XX P-P-SDB; AAB97823.
XX
PT New isolated polynucleotide, encoding spastin gene, and polypeptides,
PT useful for diagnosing autosomal recessive spastic ataxia of
PT Charlevoix-Saguenay disease by detecting two point mutations in spastin
PT gene sequence -
XX
PS Claim 1; Page -: 76pp; English.
XX
CC The present invention describes human and mouse spastin, and mutated
CC human spastin (autosomal recessive spastic ataxia of Charlevoix-Saguenay
CC (ARSACS)) gene sequences (I). The spastin gene has been mapped to
CC chromosome 13q11. (I) have neuroprotective activities and can be used in
CC gene therapy and as a spastin polypeptide agonists. (I), their fragments
CC or their complements can be useful for assaying the presence of a nucleic
CC acid molecule in a sample. (I) is useful for diagnosing or aiding in the
CC diagnosis of an early onset neurodegenerative disease in an individual.
CC The neurodegenerative disease comprises reduced sensory nerve conduction,
CC reduced motor nerve velocity, hypermyelination of retinal nerve fibres,
CC atrophy of upper cerebellar vermis, absence of Purkinje cells and
CC abnormal neuronal lipid storage. (I) can also be used to produce
CC antisense nucleic acids, is useful as molecular weight or chromosome
CC markers, to identify genetic disorders, as hybridisation probes or
CC primers, as an antigen, identify and express recombinant protein for
CC analysis, characterisation or therapeutic use, or as markers for tissues
CC in which the corresponding protein is expressed. Diagnostic methods from
CC the present invention can be used to identify subjects having or at risk
CC of developing a disease or disorder associated with aberrant expression
CC or activity of (I). The assays can be utilised to identify a subject
CC having or at risk of developing a disorder associated with spastin
CC protein or spastin gene expression or activity. The present sequence
CC encodes a mutated human spastin.
CC N.B. The present sequence is not given in the present specification but
CC is derived from the human spastin nucleotide sequence (AAH20174) as
CC stated on page 14.
XX
SQ Sequence 12793 BP; 4163 A; 2257 C; 2487 G; 3886 T; 0 other;

Query Match 92.4%; Score 1295; DB 22; Length 12793;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1295; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gtagcagtaaaactaggagcagctcccaaacgacacaaaccccttagaagaatagctcc 60
|||||
Db 5300 gtagcagtaaaactaggagcagctcccaaacgacacaaaccccttagaagaatagctcc 5359

QY 61 aatgctgtttaaacaactgtgcagaaatttggcagaagaagaataatgaccagaca 120
|||||
Db 5360 aatgctgtttaaacaactgtgcagaaatttggcagaagaagaataatgaccagaca 5419
QY 121 attaaagacatccttaatgcatatcctcttgaagaagaatgttgaagcttctcaa 180
|||||
Db 5420 attaaagacatccttaatgcatatcctcttgaagaagaatgttgaagcttctcaa 5479
QY 181 aatgctgtatgtgaagaaggcagcaagaatctgttttggtttgatccctagaagatcca 240
|||||
Db 5480 aatgctgtatgtgaagaaggcagcaagaatctgttttggtttgatccctagaagatcca 5539
QY 241 gtctagaataatttgaataaagtgagcccatgtgcagaggccagacattgtgtgtac 300
|||||
Db 5540 gtctagaataatttgaataaagtgagcccatgtgcagaggccagacattgtgtgtac 5599
QY 301 aacaacagacatttacaagaagatgattgtaggaatttcagaattcttgaagaagcagc 360
|||||
Db 5600 aacaacagacatttacaagaagatgattgtaggaatttcagaattcttgaagaagcagc 5659
QY 361 aaagaaggaaatccttataaactggaacagtatggaataagatccaattctgttatcat 420
|||||
Db 5660 aaagaaggaaatccttataaactggaacagtatggaataagatccaattctgttatcat 5719
QY 421 atccagaactgccatcttatttctgtgcaatgacatcctgttattttgatcccat 480
|||||
Db 5720 atccagaactgccatcttatttctgtgcaatgacatcctgttattttgatcccat 5779
QY 481 gccagatatgcacaaaggcgacacatccatagtcgccgagcagatgttagagatttgat 540
|||||
Db 5780 gccagatatgcacaaaggcgacacatccatagtcgccgagcagatgttagagatttgat 5839
QY 541 gcagattttggaacacagtttccagatggttctgagatctttatctggggagaccatttaa 600
|||||
Db 5840 gcagattttggaacacagtttccagatggttctgagatctttatctggggagaccatttaa 5899
QY 601 ctgataattgcacaatggttcagatttccctctctgtaatgagaagaatggcaaaagtctcg 660
|||||
Db 5900 ctgataattgcacaatggttccagatttccctctctgtaatgagaagaatggcaaaagtctcg 5959
QY 661 gaaattctgtctgttccagcatcagacagaaatggtccagaatcttcttggacaaactgyc 720
|||||
Db 5960 gaaattctgtctgttccagcatcagacagaaatggtccagaatcttcttggacaaactgyc 6019
QY 721 tcagatggggcagaactcttaatttcttaatacacaatggaataaattctcatctgtgaa 780
|||||
Db 6020 tcagatggggcagaactcttaatttcttaatacacaatggaataaattctcatctgtgaa 6079
QY 781 atagataagatgactggagctctaaatgtgtctatctcagtaaaaggcaaatccagaat 840
|||||
Db 6080 atagataagatgactggagctctaaatgtgtctatctcagtaaaaggcaaatccagaat 6139
QY 841 ggaagacagattgaaaggaaacaaattcattgcatctgtaattgtatgtagtactaaag 900
|||||
Db 6140 ggaagacagattgaaaggaaacaaattcattgcatctgtaattgtatgtagtactaaag 6199
QY 901 aggaagctcaaaagacatacccggttcaacaataactatactatgatactcgaggactct 960
|||||
Db 6200 aggaagctcaaaagacatacccggttcaacaataactatactatgatactcgaggactct 6259
QY 961 gaaagaaacttactaactgtgcttaattgtatagatcagcctttcaagatgagaagaa 1020
|||||
Db 6260 gaaagaaacttactaactgtgcttaattgtatagatcagcctttcaagatgagaagaa 6319
QY 1021 gtatctaaagtgatcatatcagctcacaagaacaaagatattactctttccacgltgt 1080
|||||
Db 6320 gtatctaaagtgatcatatcagctcacaagaacaaagatattactctttccacgltgt 6379
QY 1081 ggagtagctgctgctattaccacaactataaaaaaaccctataggcctctgtttttgt 1140
|||||
Db 6380 ggagtagctgctgctattaccacaactataaaaaaaccctataggcctctgtttttgt 6439
QY 1141 cctcttcttggagactggtgctgcatcttcatgtgaatgaccattgacatgattca 1200

|||||
Db 6440 cctcttcttggagactggtgctgcatcttcatgtgaatgaccattgacatgattca 6499
QY 1201 gccagaaggaacctgtgctgcatgataatgagttgtgttcgaagtgaactggaatac 1260
|||||
Db 6500 gccagaaggaacctgtgctgcatgataatgagttgtgttcgaagtgaactggaatac 6559
QY 1261 agttaatgacagcatatagctcttcataatgt 1295
|||||
Db 6560 agttaatgacagcatatagctcttcataatgt 6594
RESULT 6
AAH20175
ID AAH20175 standard; DNA; 11493 BP.
XX
AC AAH20175;
XX
DT 09-AUG-2001 (first entry)
XX
DE Mouse spastin nucleotide sequence SEQ ID NO:3.
XX
XX Human; mouse; spastin; ARSACS; chromosome 13q11; identification;
KW autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;
KW neurodegenerative disease; reduced sensory nerve conduction; diagnosis;
KW reduced motor nerve velocity; hypermyelination of retinal nerve fibre;
KW atrophy of upper cerebellar vermis; absence of Purkinje cell;
KW abnormal neuronal lipid storage; genetic disorder; characterisation; ds.
OS Mus musculus.
XX
XX
FH Key Location/Qualifiers
FT 1..11493
FT CDS /tag="a
FT /product="spastin"
XX
PN MO200129266-A2.
XX
XX 26-APR-2001.
PD
XX 20-OCT-2000; 2000WO-US29130.
PE
XX 20-OCT-1999; 99US-0160588.
PR
XX
PA (UTMC-) UNIV MCGILL.
PA (HOPI-) HOPITAL SAINTE-JUSTINE.
PI Hudson TJ, Engert J, Richter A;
XX
DR WPI: 2001-308494/32.
DR P-PSDB: AAB97820.
XX
PT New isolated polynucleotide, encoding spastin gene, and polypeptides,
PT useful for diagnosing autosomal recessive spastic ataxia of
PT Charlevoix-Saguenay disease by detecting two point mutations in spastin
PT gene sequence -
XX
PS Claim 1: Fig 8; 76pp; English.
XX
CC The present invention describes human and mouse spastin, and mutated
CC human spastin (autosomal recessive spastic ataxia of Charlevoix-Saguenay
CC (ARSACS)) gene sequences (I). The spastin gene has been mapped to
CC chromosome 13q11. (I) have neuroprotective activities and can be used in
CC gene therapy and as a spastin polypeptide agonists. (I), their fragments
CC or their complements can be useful for assaying the presence of a nucleic
CC acid molecule in a sample. (I) is useful for diagnosing or aiding in the
CC diagnosis of an early onset neurodegenerative disease in an individual.
CC The neurodegenerative disease comprises reduced sensory nerve conduction,
CC reduced motor nerve velocity, hypermyelination of retinal nerve fibres,
CC atrophy of upper cerebellar vermis, absence of Purkinje cells and
CC abnormal neuronal lipid storage. (I) can also be used to produce
CC antisense nucleic acids, is useful as molecular weight or chromosome
CC markers, to identify genetic disorders, as hybridisation probes or

CC primers, as an antigen, identify and express recombinant protein for
CC analysis, characterisation or therapeutic use, or as markers for tissues
CC in which the corresponding protein is expressed. Diagnostic methods from
CC the present invention can be used to identify subjects having or at risk
CC of developing a disease or disorder associated with aberrant expression
CC or activity of (I). The assays can be utilised to identify a subject
CC having or at risk of developing a disorder associated with spastin
CC protein or spastin gene expression or activity. The present sequence
CC encodes mouse spastin as given in the present invention.

SQ Sequence 11493 BP: 3599 A; 2281 C; 2387 G; 3226 T; 0 other:

Query Match 3.8%; Score 53; DB 22; Length 11493;

Best Local Similarity 100.0%; Pred. No. 8.5e-16;

Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 298 tacacacaccgaccattacagaagatgatgttagaggaattcagaactcttg 350

DB 5521 tacacacaccgaccattacagaagatgatgttagaggaattcagaactcttg 5573

RESULT 7

AAH20151/c

AAH20151 standard; DNA: 24 BP.

AC AAH20151;

DT 09-AUG-2001 (first entry)

DE Human spastin ORF PCR primer SEQ ID NO:44.

KM Human: mouse; spastin; ARSACS; chromosome 13q11; identification;

KM autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;

KM neurodegenerative disease; reduced sensory nerve conduction; diagnosis;

KM reduced motor nerve velocity; hypermyelination of retinal nerve fibre;

KM atrophy of upper cerebellar vermis; absence of Purkinje cell;

KM abnormal neuronal lipid storage; genetic disorder; characterisation;

KM PCR primer; ss.

XX Homo sapiens.

OS Synthetic.

PN WO200129266-A2.

PA (UYMC-) UNIV MCGILL.

PA (HOP1-) HOPITAL SAINTE-JUSTINE.

PI Hudson TJ, Engert J, Richter A;

DR WPI: 2001-308494/32.

PT New isolated polynucleotide, encoding spastin gene, and polypeptides,
PT useful for diagnosing autosomal recessive spastic ataxia of
PT Charlevoix-Saguenay disease by detecting two point mutations in spastin
PT gene sequence -

PS Claim 23; Fig 7; 76pp: English.

CC The present invention describes human and mouse spastin, and mutated
CC human spastin (autosomal recessive spastic ataxia of Charlevoix-Saguenay
CC (ARSACS)) gene sequences (I). The spastin gene has been mapped to
CC chromosome 13q11. (I) have neuroprotective activities and can be used in
CC gene therapy and as a spastin polypeptide agonists. (I), their fragments
CC or their complements can be useful for assaying the presence of a nucleic
CC acid molecule in a sample. (I) is useful for diagnosing or aiding in the
CC diagnosis of an early onset neurodegenerative disease in an individual.

CC The neurodegenerative disease comprises reduced sensory nerve conduction,
CC reduced motor nerve velocity, hypermyelination of retinal nerve fibres,
CC atrophy of upper cerebellar vermis, absence of Purkinje cells and
CC abnormal neuronal lipid storage. (I) can also be used to produce
CC antisense nucleic acids, is useful as molecular weight or chromosome
CC markers, to identify genetic disorders, as hybridisation probes or
CC primers, as an antigen, identify and express recombinant protein for
CC analysis, characterisation or therapeutic use, or as markers for tissues
CC in which the corresponding protein is expressed. Diagnostic methods from
CC the present invention can be used to identify subjects having or at risk
CC of developing a disease or disorder associated with aberrant expression
CC or activity of (I). The assays can be utilised to identify a subject
CC having or at risk of developing a disorder associated with spastin
CC protein or spastin gene expression or activity. AAH20122 to AAH20173
CC represent specifically claimed primers which can be used in diagnostic
CC methods from the present invention.

SQ Sequence 24 BP: 8 A; 6 C; 4 G; 6 T; 0 other:

Query Match 1.7%; Score 24; DB 22; Length 24;

Best Local Similarity 100.0%; Pred. No. 0.16; Mismatches 0; Indels 0; Gaps 0;

QY 793 actgagctctaaatgtgctgtat 816

DB 24 ACTGAGCTCTAAATGTCGTGTA 1

RESULT 8

AAH20149/c

AAH20149 standard; DNA: 21 BP.

AC AAH20149;

DT 09-AUG-2001 (first entry)

DE Human spastin ORF PCR primer SEQ ID NO:42.

KM Human: mouse; spastin; ARSACS; chromosome 13q11; identification;

KM autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;

KM neurodegenerative disease; reduced sensory nerve conduction; diagnosis;

KM reduced motor nerve velocity; hypermyelination of retinal nerve fibre;

KM atrophy of upper cerebellar vermis; absence of Purkinje cell;

KM abnormal neuronal lipid storage; genetic disorder; characterisation;

KM PCR primer; ss.

XX Homo sapiens.

OS Synthetic.

PN WO200129266-A2.

PA (UYMC-) UNIV MCGILL.

PA (HOP1-) HOPITAL SAINTE-JUSTINE.

PI Hudson TJ, Engert J, Richter A;

DR WPI: 2001-308494/32.

PT New isolated polynucleotide, encoding spastin gene, and polypeptides,
PT useful for diagnosing autosomal recessive spastic ataxia of
PT Charlevoix-Saguenay disease by detecting two point mutations in spastin
PT gene sequence -

PS Claim 23; Fig 7; 76pp: English.

CC The present invention describes human and mouse spastin, and mutated

human spastin (autosomal recessive spastic ataxia of Charlevoix-Saguenay (ASACS)) gene sequences (I). The spastin gene has been mapped to chromosome 1q11. (I) have neuroprotective activities and can be used in gene therapy and as a spastin polypeptide agonists. (I), their fragments or their complements can be useful for assaying the presence of a nucleic acid molecule in a sample. (I) is useful for diagnosing or aiding in the diagnosis of an early onset neurodegenerative disease in an individual. The neurodegenerative disease comprises reduced sensory nerve conduction, reduced motor nerve velocity, hypermyelination of retinal nerve fibres, atrophy of upper cerebellar vermis, absence of Purkinje cells and abnormal neuronal lipid storage. (I) can also be used to produce antisense nucleic acids, is useful as molecular weight or chromosome markers, to identify genetic disorders, as hybridisation probes or primers, as an antigen, identify and express recombinant protein for analysis, characterisation or therapeutic use, or as markers for tissues in which the corresponding protein is expressed. Diagnostic methods from the present invention can be used to identify subjects having or at risk of developing a disease or disorder associated with aberrant expression or activity of (I). The assays can be utilised to identify a subject having or at risk of developing a disorder associated with Spastin protein or spastin gene expression or activity. AAH02122 to AAH02173 represent specifically claimed primers which can be used in diagnostic methods from the present invention.

Sequence 21 BP; 5 A; 7 C; 5 G; 4 T; 0 other;

Query Match	1.5%	Score 21;	DB 22;	length 21;
Best Local Similarity	100.0%	Pred. No. 4.7;		
Matches 21; Conservative	0;	Mismatches	0;	Indels 0;
				Gaps 0;

```
QY      276 gcaaggccagcacttgtgt 296
          |||||
Db      21  GCAAGGCCAGCAGCTTGTGT 1
```

RESULT	9
AAH20150	
ID	AAH20150 standard; DNA; 21 BP.

AC AAH20150;

DT 09-AUG-2001 (first entry)

DE Human spastin ORF PCR primer SEQ ID NO:43.

Human mouse: spstn1: ARSACS: chromosome 13q11: identification:
 KW autosomal recessive spastic ataxia of Charlevoix-Saguenay: mutation:
 KW neurodegenerative disease; reduced sensory nerve conduction; diagnosis
 KW reduced motor nerve velocity; hypermyelination of retinal nerve fibre;
 KW atrophy of upper cerebellar vermis; absence of Purkinje cell;
 KW abnormal neuronal lipid storage; genetic disorder; characterisation;
 KW PCR primer: ss.

OS Homo sapiens.
OS Synthetic.

PN W0200129266-A2.

PD 26-APR-2001.

PF 20-OCT-2000; 2000WO-US29130.

PR 20-OCT-1999; 99US-0160588.

PA (UYMC-) UNIV MCGILL.
PA (HOPI-) HOPITAL SAINTE-JUSTINE.

PI Hudson TJ, Engert J, Richter A;

DR WPI; 2001-308494/32.

PT New isolated polynucleotide, encoding spastin gene, and polypeptides,

PT useful for diagnosing autosomal recessive spastic ataxia of
PT Charlevoix-Saguenay disease by detecting two point mutations in spastin
PT gene sequence -

PS Claim 23; Fig 7; 76pp; English.

CC The present invention describes human and mouse spastin, and mutated
CC human spastin (autosomal recessive spastic ataxia of Charlevoix-Saguenay
CC (ARSACS)) gene sequences (1). The spastin gene has been mapped to
CC chromosome 13q11. (1) have neuroprotective activities and can be used in
CC gene therapy and as a spastin polypeptide agonists. (1), their fragments
CC or their complements can be useful for assaying the presence of a nucleic
CC acid molecule in a sample. (1) is useful for diagnosing or aiding in the
CC diagnosis of an early onset neurodegenerative disease in an individual.
CC The neurodegenerative disease comprises reduced sensory nerve conduction,
CC reduced motor nerve velocity, hypermyelination of retinal nerve fibres,
CC atrophy of upper cerebellar vermis, absence of Purkinje cells and
CC abnormal neuronal lipid storage. (1) can also be used to produce
CC antisense nucleic acids, is useful as molecular weight or chromosome
CC markers, to identify genetic disorders, as hybridisation probes or
CC primers, as an antigen, identify and express recombinant protein for
CC analysts, characterisation or therapeutic use, or as markers for tissues
CC in which the corresponding protein is expressed. Diagnostic methods from
CC the present invention can be used to identify subjects having or at risk
CC of developing a disease or disorder associated with aberrant expression
CC or activity of (1). The assays can be utilised to identify a subject
CC having or at risk of developing a disorder associated with Spastin
CC protein or spastin gene expression or activity. AAH20122 to AAH20173
CC represent specifically claimed primers which can be used in diagnostic
CC methods from the present invention.

Sequence 21 BP; 9 A; 4 C; 6 G; 2 T; 0 other;

Query Match	1.5%	Score 21	DB 22	length 21
Best Local Similarity	100.0%	Pred. No. 4.7		
Matches 21, Conservative	0	Mismatches	0	Indels 0; Gaps 0;

QY 190 gatgcaagcgacagaatc 210

Db 1 gatgcaagcgacagaatc 21

RESULT 10

ID AAH20152 standard; DNA; 21 BP.

AC AAH20152;

DT 09-AUG-2001 (first entry)

	Human spastin ORF PCR primer SEQ ID NO:45.
DE	

KW Human; onuse; spstn; ARSACS; chromosome 1q11; identification;
 KW autosomal recessive spastic ataxia of Charlevoix-Seguinay; mutation;
 KW neurodegenerative disease; reduced sensory nerve conduction; diagnosis;
 KW reduced motor nerve velocity; hypermyelination of retinal nerve fibre;
 KW atrophy of upper cerebellar vermis; absence of Purkinje cell;
 KW abnormal neuronal lipid storage; genetic disorder; characterisation;
 KW PCR primer; ss.

OS Homo sapiens.

XX

XX

PF 20-OCT-2000; 2000WO-US29130.

PR 20-OCT-1999; 99US-0160588.

PA
(UYMC-) UNIV MCGILL.

PA (HOP1-) HOPITAL SAINTE-JUSTINE.
 XX
 PI Hudson TJ, Engert J, Richter A;
 XX
 DR WPI; 2001-308494/32.
 XX
 PT New isolated polynucleotide, encoding spastin gene, and polypeptides,
 PT useful for diagnosing autosomal recessive spastic ataxia of
 PT Charlevoix-Saguenay disease by detecting two point mutations in spastin
 PT gene sequence -
 XX
 PI Hudson TJ, Richter A;
 XX
 PS WPI; 2001-308494/32.
 XX
 CC The present invention describes human and mouse spastin, and mutated
 CC human spastin (autosomal recessive spastic ataxia of Charlevoix-Saguenay
 CC (ARSCS)) gene sequences (I). The spastin gene has been mapped to
 CC chromosome 13q11. (I) have neuroprotective activities and can be used in
 CC gene therapy and as a spastin polypeptide agonists. (I), their fragments
 CC or their complements can be useful for assaying the presence of a nucleic
 CC acid molecule in a sample. (I) is useful for diagnosing or aiding in the
 CC diagnosis of an early onset neurodegenerative disease in an individual.
 CC The neurodegenerative disease comprises reduced sensory nerve conduction,
 CC reduced motor nerve velocity, hypermyelination of retinal nerve fibres,
 CC atrophy of upper cerebellar vermis, absence of Purkinje cells and
 CC abnormal neuronal lipid storage. (I) can also be used to produce
 CC antisense nucleic acids, is useful as molecular weight or chromosome
 CC markers, to identify genetic disorders, as hybridisation probes or
 CC primers, as an antigen, identify and express recombinant protein for
 CC analysis, characterisation or therapeutic use, or as markers for tissues
 CC in which the corresponding protein is expressed. Diagnostic methods from
 CC the present invention can be used to identify subjects having or at risk
 CC of developing a disease or disorder associated with aberrant expression
 CC or activity of (I). The assays can be utilised to identify a subject
 CC having or at risk of developing a disorder associated with spastin
 CC protein or spastin gene expression or activity. AAH20122 to AAH20173
 CC represent specifically claimed primers which can be used in diagnostic
 CC methods from the present invention.
 CC
 XX
 SQ Sequence 21 BP; 7 A; 5 C; 6 G; 3 T; 0 other;
 XX
 QY Query Match 1.5%; Score 21; DB 22; Length 21;
 Best Local Similarity 100.0%; Pred. No. 4.7;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 XX
 Db 679 gcatcagacagatgctccag 699
 1 gcatcagacagatgctccag 21
 XX
 RESULT 11
 AAH20154
 ID AAH20154 standard; DNA; 20 BP.
 XX
 AC AAH20154;
 XX
 DT 09-AUG-2001 (first entry)
 XX
 DE Human spastin ORF PCR primer SEQ ID NO:47.
 XX
 KW Human; mouse; spastin; ARSCS; chromosome 13q11; identification;
 KW autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;
 KW neurodegenerative disease; reduced sensory nerve conduction; diagnosis;
 KW reduced motor nerve velocity; hypermyelination of retinal nerve fibre;
 KW atrophy of upper cerebellar vermis; absence of Purkinje cell;
 KW abnormal neuronal lipid storage; genetic disorder; characterisation;
 KW PCR primer; ss.
 XX
 OS Homo sapiens.
 OS Synthetic.
 XX
 WO200129266-A2.
 XX

PD 26-APR-2001.
 XX
 PF 20-OCT-2000; 2000MO-US29130.
 XX
 PR 20-OCT-1999; 99US-0160588.
 XX
 PA (UYMC-) UNIV MCGILL.
 PA (HOP1-) HOPITAL SAINTE-JUSTINE.
 PI Hudson TJ, Engert J, Richter A;
 XX
 DR WPI; 2001-308494/32.
 XX
 PT New isolated polynucleotide, encoding spastin gene, and polypeptides,
 PT useful for diagnosing autosomal recessive spastic ataxia of
 PT Charlevoix-Saguenay disease by detecting two point mutations in spastin
 PT gene sequence -
 XX
 PI Hudson TJ, Richter A;
 XX
 PS WPI; 2001-308494/32.
 XX
 CC The present invention describes human and mouse spastin, and mutated
 CC human spastin (autosomal recessive spastic ataxia of Charlevoix-Saguenay
 CC (ARSCS)) gene sequences (I). The spastin gene has been mapped to
 CC chromosome 13q11. (I) have neuroprotective activities and can be used in
 CC gene therapy and as a spastin polypeptide agonists. (I), their fragments
 CC or their complements can be useful for assaying the presence of a nucleic
 CC acid molecule in a sample. (I) is useful for diagnosing or aiding in the
 CC diagnosis of an early onset neurodegenerative disease in an individual.
 CC The neurodegenerative disease comprises reduced sensory nerve conduction,
 CC reduced motor nerve velocity, hypermyelination of retinal nerve fibres,
 CC atrophy of upper cerebellar vermis, absence of Purkinje cells and
 CC abnormal neuronal lipid storage. (I) can also be used to produce
 CC antisense nucleic acids, is useful as molecular weight or chromosome
 CC markers, to identify genetic disorders, as hybridisation probes or
 CC primers, as an antigen, identify and express recombinant protein for
 CC analysis, characterisation or therapeutic use, or as markers for tissues
 CC in which the corresponding protein is expressed. Diagnostic methods from
 CC the present invention can be used to identify subjects having or at risk
 CC of developing a disease or disorder associated with aberrant expression
 CC or activity of (I). The assays can be utilised to identify a subject
 CC having or at risk of developing a disorder associated with spastin
 CC protein or spastin gene expression or activity. AAH20122 to AAH20173
 CC represent specifically claimed primers which can be used in diagnostic
 CC methods from the present invention.
 CC
 XX
 SQ Sequence 20 BP; 4 A; 5 C; 5 G; 6 T; 0 other;
 XX
 QY Query Match 1.4%; Score 20; DB 22; Length 20;
 Best Local Similarity 100.0%; Pred. No. 15;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 XX
 Db 1174 gtgaatgccaacttgcact 1193
 1 gtgaatgccaacttgcact 20
 XX
 RESULT 12
 ABL25800/C
 ID ABL25800 standard; DNA; 8496 BP.
 XX
 AC ABL25800;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 28873.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ds.
 XX
 OS Drosophila melanogaster.
 XX
 WO200171042-A2.
 XX

XX 27-SEP-2001.
PD
XX
XX 23-MAR-2001; 2001WO-US09231.
PE
XX 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PMD, Myers EW;
DR WPI; 2001-656860/75.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX
PS Claim 1; SEQ ID NO 28873; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-AB16351), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins
CC (AB857737-AB872072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 8496 BP; 2387 A; 1826 C; 1971 G; 2312 T; 0 other;

Query Match 1.4%; Score 20; DB 23; Length 8496;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 802 ctaaatgctgtatcagt 821
|||||
Db 3486 CTAATGTGCTGTATTCAGT 3467

RESULT 13
ID AAF66282/c
XX AAF66282 standard; cDNA; 389 BP.
XX
AC AAF66282;
XX
DT 09-APR-2001 (first entry)
XX
XX Novel human polynucleotide, SEQ ID NO: 2038.
XX
XX Human; cytostatic; gene therapy: colon cancer; prostate cancer;
KM breast cancer; lung cancer; cancer detection; ss.
XX
XX Homo sapiens.
OS
XX WO200102568-A2.
PN
XX
PD 11-JAN-2001.
XX
XX 30-JUN-2000; 2000WO-US18374.
XX
XX 02-JUL-1999; 99US-0142310.
PR 02-JUL-1999; 99US-0142311.
XX
XX (CHIR) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX
PI Williams LT, Escobedo J, Innis MA, Garcia PD, Klinger J, Kassam A;
PI Reinhard C, Randazzo F, Kennedy GC, Pot D, Lamson G, Drmanac R;

PI Crkenjakov R, Drmanac S, Dickson M, Labat I, Leschkowitz D;
PI Kita D, Garcia V, Jones LM, Strache-Crain B;
XX WPI; 2001-091805/10.
DR
XX
XX Library of polynucleotides for diagnosing a cancerous state of a
PT mammalian cell and detecting cancer, particularly of the colon or
PT prostate, comprises 3351 human polynucleotide sequences -
XX
XX
PS Claim 9; Page 839; 1046pp; English.
XX
XX The present sequence is one of 3351 sequences in a library of human
CC polynucleotides. The library is used to detect differentially expressed
CC genes correlated with a cancerous state of a mammalian cell and can
CC detect colon, prostate, breast and lung cancer. The library can be used
CC to produce probes for detection of mRNA and to produce additional copies
CC of the polynucleotides. The probes can be used for chromosome mapping of
CC the polynucleotide and for detection of transcription levels. Ribozymes
CC or antisense oligonucleotides can be generated. The polynucleotides and
CC their gene products are used as genetic or biochemical markers (e.g. in
CC blood or tissues) that will detect the earliest changes along the
CC carcinogenesis pathway and/or monitor the efficacy of therapies and
CC preventive interventions. The polynucleotides, polypeptides and
CC antibodies against them can be used in pharmaceutical compositions to
CC treat the cancers and proliferative disorders such as neoplasia,
CC dysplasia and hyperplasia.
XX
SQ Sequence 389 BP; 107 A; 76 C; 83 G; 123 T; 0 other;

Query Match 1.4%; Score 19; DB 22; Length 389;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 402 attcaattctgtatcat 420
|||||
Db 215 ATTCAATCTGTGTATCAT 197

RESULT 14
ID AAS83961/c
XX AAS83961 standard; cDNA; 448 BP.
XX
AC AAS83961;
XX
DT 13-FEB-2002 (first entry)
XX
XX DNA encoding novel human diagnostic protein #19765.
DE
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KM food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX Homo sapiens.
OS
XX WO200175067-A2.
PN
XX
PD 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
PI WPI; 2001-639362/73.
DR P-PSDB; ABG19774.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 22, 2002, 08:29:03 ; Search time 91.58 Seconds
(without alignments)
3757.726 Million cell updates/sec

Title: US-09-693-205-7_COPY_5300_6700

Perfect score: 1401
Sequence: 1 gtacgagcaaaactagagc.....aaagacacttaagaagt 1401

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 383533 seqs, 122816752 residues

Word size : 0

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

- Issued_Patents_NA:*
- 1: /cgn2_6/prodata/1/ina/5A_COMB.seq:*
 - 2: /cgn2_6/prodata/1/ina/5B_COMB.seq:*
 - 3: /cgn2_6/prodata/1/ina/6A_COMB.seq:*
 - 4: /cgn2_6/prodata/1/ina/6B_COMB.seq:*
 - 5: /cgn2_6/prodata/1/ina/PCBUS_COMB.seq:*
 - 6: /cgn2_6/prodata/1/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	18	1.3	1806	2	US-08-819-013-2
2	17	1.2	2157	1	US-08-451-715A-3
3	17	1.2	2837	4	US-09-156-316-11
4	17	1.2	176373	3	US-09-128-155-17
5	16	1.1	20	4	US-09-467-642-57
6	16	1.1	269	3	US-08-258-287B-12
7	16	1.1	270	3	US-08-368-704C-12
8	16	1.1	345	4	US-08-391-789A-224
9	16	1.1	345	4	US-09-062-451-224
10	16	1.1	468	5	PCT-US91-08177-8
11	16	1.1	777	2	US-08-975-316-37
12	16	1.1	1299	1	US-07-688-352C-17
13	16	1.1	1299	2	US-08-474-379C-17
14	16	1.1	1299	3	US-09-146-249A-17
15	16	1.1	1299	3	US-08-206-188B-17
16	16	1.1	1299	5	PCT-US91-02714-11
17	16	1.1	1299	5	PCT-US91-02714-17
18	16	1.1	1341	4	US-09-018-635-30
19	16	1.1	1368	3	US-09-018-628-15
20	16	1.1	1368	3	US-09-273-378-15
21	16	1.1	1415	3	US-08-413-118-127
22	16	1.1	1415	3	US-08-473-446-127
23	16	1.1	1486	3	US-08-956-182-24
24	16	1.1	1511	3	US-08-956-182-41
25	16	1.1	1732	4	US-09-449-335-1
26	16	1.1	1732	4	US-09-449-335-5
27	16	1.1	1951	1	US-08-487-890A-112

28	16	1.1	1951	2	US-08-478-435-112	Sequence 112, App
29	16	1.1	1951	2	US-08-337-483-112	Sequence 112, App
30	16	1.1	1951	2	US-08-478-373-112	Sequence 112, App
31	16	1.1	1951	3	US-08-474-671-112	Sequence 112, App
32	16	1.1	1951	3	US-08-483-577A-112	Sequence 112, App
33	16	1.1	1951	3	US-08-897-438-112	Sequence 112, App
34	16	1.1	1951	4	US-08-637-654-112	Sequence 112, App
35	16	1.1	2057	3	US-09-008-103-1	Sequence 1, App1
36	16	1.1	2089	4	US-09-155-770-6	Sequence 6, App1
37	16	1.1	2119	4	US-09-018-635-28	Sequence 28, App1
38	16	1.1	2375	1	US-08-369-796-9	Sequence 9, App1
39	16	1.1	2375	2	US-08-852-091-9	Sequence 9, App1
40	16	1.1	2375	2	US-08-820-754-9	Sequence 9, App1
41	16	1.1	2375	3	US-08-956-652-9	Sequence 9, App1
42	16	1.1	2375	3	US-08-956-869-9	Sequence 9, App1
43	16	1.1	2375	3	US-08-948-547-9	Sequence 9, App1
44	16	1.1	2375	5	PCT-US95-17025-9	Sequence 9, App1
45	16	1.1	2525	4	US-09-309-487-24	Sequence 24, App1

ALIGNMENTS

RESULT 1
; Sequence 2, Application US/08819013
; Patent No. 5994522
; GENERAL INFORMATION:
; APPLICANT: Chan, Andrew C.
; TITLE OF INVENTION: BLNK PROTEINS
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Honbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/819,013
; FILING DATE: 17-MAR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/788,322
; FILING DATE: 24-JAN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: A-64383-1/RFT/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEO ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1806 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA
; US-08-819-013-2

Query Match 1.38; Score 18; DB 2; Length 1806;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 92 ttggcgagaagaagaat 109

Db 1378 TTGGCGAAGAAAGAAAT 1395

RESULT 2

US-08-451-715A-3
Sequence 3, Application US/08451715A
Patent No. 5801013
GENERAL INFORMATION:
APPLICANT: Tao, Jianshi
APPLICANT: Qiu, Yan
APPLICANT: Houman, Fariba
APPLICANT: Shen, Xiaoyu
TITLE OF INVENTION: Helicobacter Aminoacyl-tRNA Synthetase
TITLE OF INVENTION: Proteins, Nucleic Acids and Strains Comprising Same
NUMBER OF SEQUENCES: 67
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Millitia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: USA
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/451,715A
FILING DATE: 26-MAY-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Brook, David E.
REGISTRATION NUMBER: 22,592
REFERENCE/DOCKET NUMBER: CPT94-25
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-861-6240
TELEFAX: 617-861-9540
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2157 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 102..2045
US-08-451-715A-3

Query Match 1.2%; Score 17; DB 1; Length 2157;

Best Local Similarity 100.0%; Pred. No. 36;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1385 ggacacttaagaagt 1401

Db 185 GGACACTTTAAGAACT 201

RESULT 3

US-09-156-316-11
Sequence 11, Application US/09156316
Patent No. 6183961
GENERAL INFORMATION:
APPLICANT: Bernstein, Harold S.
APPLICANT: Coughlin, Shaun R.
TITLE OF INVENTION: Methods and Compositions for Regulating Cell Cycle
FILE REFERENCE: UCSF-020/01US
CURRENT APPLICATION NUMBER: US/09/156,316
CURRENT FILING DATE: 1998-09-18

EARLIER APPLICATION NUMBER: 60/060,688
EARLIER FILING DATE: 1997-09-22
NUMBER OF SEQ ID NOS: 12
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 11
LENGTH: 2837
TYPE: DNA
ORGANISM: Homo sapiens
US-09-156-316-11

Query Match

1.2%; Score 17; DB 4; Length 2837;

Best Local Similarity 100.0%; Pred. No. 36;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 684 agacagaatggtccaga 700

Db 280 agacagaatggtccaga 296

RESULT 4

US-09-128-155-17/C
Sequence 17, Application US/09128155
Patent No. 6117654
GENERAL INFORMATION:
APPLICANT: Pan, Yang
TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
FILE REFERENCE: 09404/052001
CURRENT APPLICATION NUMBER: US/09/128,155
CURRENT FILING DATE: 1998-08-03
EARLIER APPLICATION NUMBER: US 60/091,650
EARLIER FILING DATE: 1998-07-02
EARLIER APPLICATION NUMBER: US 60/054,646
EARLIER FILING DATE: 1997-08-04
NUMBER OF SEQ ID NOS: 18
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 17
LENGTH: 176373
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(176373)
OTHER INFORMATION: n = A,T,C or G
US-09-128-155-17

Query Match

1.2%; Score 17; DB 3; Length 176373;

Best Local Similarity 100.0%; Pred. No. 35;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 751 aatcacatggaanaat 767

Db 102233 AATCACATGGAANAAT 102217

RESULT 5

US-09-467-642-57/C
Sequence 57, Application US/09467642
Patent No. 6300132
GENERAL INFORMATION:
APPLICANT: Brett P. Monia
APPLICANT: lex M. Cowart
TITLE OF INVENTION: ANTISENSE MODULATION OF TELOMERIC REPEAT BINDING FACTOR 2 EXP
CURRENT APPLICATION NUMBER: US/09/467,642
CURRENT FILING DATE: 1999-12-20
NUMBER OF SEQ ID NOS: 89
SEQ ID NO 57
LENGTH: 20
TYPE: DNA
ORGANISM: Artificial Sequence

FEATURE:
OTHER INFORMATION: Antisense Oligonucleotide
US-09-467-642-57

Query Match 1.1%; Score 16; DB 4; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1156 actgggctgccattc 1171
|||||

Db 20 ACTGGCTGCCATTTC 5

RESULT 6
US-08-258-287B-12
Sequence 12, Application US/08258287B
Patent No. 6083735
GENERAL INFORMATION:
APPLICANT: Yuan, Junying
TITLE OF INVENTION: Programmed Cell Death Genes and Proteins
NUMBER OF SEQUENCES: 85
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox
STREET: 1100 New York Avenue, Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/258,287B
FILING DATE: 10-JUN-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/080,850
FILING DATE: 24-JUN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Bugalsky, Lawrence B.
REGISTRATION NUMBER: 35,086
REFERENCE/DOCKET NUMBER: 0609.3920001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
TELEX: 248636 SSK
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 269 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: both
US-08-258-287B-12

Query Match 1.1%; Score 16; DB 3; Length 269;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1366 cccctatcatgttgt 1381
|||||

Db 168 CCCCTATTCATGTGT 183

RESULT 7
US-08-368-704C-12
Sequence 12, Application US/08368704C
Patent No. 6087160
GENERAL INFORMATION:

APPLICANT: Yuan, Junying
APPLICANT: Miura, Masayuki
TITLE OF INVENTION: Programmed Cell Death Genes and Proteins
NUMBER OF SEQUENCES: 95
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox
STREET: 1100 New York Avenue, Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/368,704C
FILING DATE: 4-JAN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/258,287
FILING DATE: 10-JUN-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/080,850
FILING DATE: 24-JUN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Bugalsky, Lawrence B.
REGISTRATION NUMBER: 35,086
REFERENCE/DOCKET NUMBER: 0609.3920002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
TELEX: 248636 SSK
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 270 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: both
US-08-368-704C-12

Query Match 1.1%; Score 16; DB 3; Length 270;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1366 cccctatcatgttgt 1381
|||||

Db 168 CCCCTATTCATGTGT 183

RESULT 8
US-08-991-789A-224/C
Sequence 224, Application US/08991789A
Patent No. 6225054
GENERAL INFORMATION:
APPLICANT: Fridakis, Tony N.
Smith, John M.
Reed, Steven G.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
TREATMENT AND DIAGNOSIS OF BREAST CANCER
NUMBER OF SEQUENCES: 292
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seed IP Law Group
STREET: 701 Fifth Avenue, Suite 6300
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/08/991,789A
APPLICATION NUMBER: US/08/991,789A
FILING DATE: 11-Dec-1997
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Potler, Jane E. R.
REGISTRATION NUMBER: 33,332
REFERENCE/DOCKET NUMBER: 210121.419C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 224:
SEQUENCE CHARACTERISTICS:
LENGTH: 345 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-991-789A-224
SEQUENCE DESCRIPTION: SEQ ID NO: 224:

Query Match 1.1%; Score 16; DB 4; Length 345;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 760 gaaaaaattctattt 775
|||||
DB 183 GAAAAAATTCATTT 168

RESULT 9
US-09-062-451-224/C
Sequence 224, Application US/09062451
Patent No. 6344550
GENERAL INFORMATION:
APPLICANT: Frudakis, Tony N.
APPLICANT: Smith, John M.
APPLICANT: Reed, Steven G.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
TREATMENT AND DIAGNOSIS OF BREAST CANCER
NUMBER OF SEQUENCES: 297
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/062,451
FILING DATE: 04-APR-1997
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.419C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 224:
SEQUENCE CHARACTERISTICS:
LENGTH: 345 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
A

US-09-062-451-224
Query Match 1.1%; Score 16; DB 4; Length 345;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 760 gaaaaaattctattt 775
|||||
DB 183 GAAAAAATTCATTT 168

RESULT 10
PCT-US91-08177-8/C
Sequence 8, Application PC/TUS9108177
GENERAL INFORMATION:
APPLICANT: Samal, Sida K
TITLE OF INVENTION: Bovine Respiratory Syncytial Virus Genes
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Venable, Baetjer, Howard & Civiletti
STREET: 1201 New York Avenue N.W., suite 1000
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US91/08177
FILING DATE: 19911104
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/608,937
FILING DATE: 05-NOV-1990
ATTORNEY/AGENT INFORMATION:
NAME: Higbet, David W
REGISTRATION NUMBER: 30,265
REFERENCE/DOCKET NUMBER: 20509-96711
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-962-4854
TELEFAX: 202-962-8300
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 468 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Bovine respiratory syncytial virus
STRAIN: A 51908
FEATURE:
NAME/KEY: CDS
LOCATION: 84..302
OTHER INFORMATION: /label= SH gene
PCT-US91-08177-8

Query Match 1.1%; Score 16; DB 5; Length 468;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 535 ttgatgcagatttta 550
|||||
DB 25 TTGATGCAGATTTTA 10

RESULT 11
US-08-975-316-37
; Sequence 37, Application US/08975316
; Patent No. 5952486
; GENERAL INFORMATION:
; APPLICANT: BLOKSBERG, Leonard N., HAVUKKALA, Ilkka
; APPLICANT: and GRIERSON, Alastair W.
; TITLE OF INVENTION: MATERIALS AND METHODS FOR
; TITLE OF INVENTION: THE MODIFICATION OF PLANT LIGNIN CONTENT
; NUMBER OF SEQUENCES: 88
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Law Offices of Ann W. Speckman
; STREET: 2601 Elliott Avenue, Suite 4185
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/975,316
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/713,000
; FILING DATE: September 11, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: SLEATH, Janet
; REGISTRATION NUMBER: 37,007
; REFERENCE/DOCKET NUMBER: 11000/1003C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-269-0565
; TELEFAX: 206-269-0563
; TELEX:
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 777 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-975-316-37

Query Match 1.1%; Score 16; DB 2; Length 777;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 751 aatcaatggaataaa 766
Db 755 AATCATGATGAAAAA 770

RESULT 12
US-07-688-352C-17
; Sequence 17, Application US/07688352C
; Patent No. 5527896
; GENERAL INFORMATION:
; APPLICANT: Wigler, Michael H.
; APPLICANT: Colicelli, John J.
; TITLE OF INVENTION: Cloning by Complementation and Related
; TITLE OF INVENTION: Processes
; NUMBER OF SEQUENCES: 57
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
; STREET: Two First National Plaza, 20 South Clark
; STREET: Street
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA

ZIP: 60603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/688,352C
; FILING DATE: 19910419
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/511,715
; FILING DATE: 20-APR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Borun, Michael F.
; REGISTRATION NUMBER: 25447
; REFERENCE/DOCKET NUMBER: 27805/30197
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 346-5750
; TELEFAX: (312) 984-9740
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1299 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1299
; US-07-688-352C-17

Query Match 1.1%; Score 16; DB 1; Length 1299;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 960 tgaaggaatcttact 975
Db 686 TGAAGGAATCTTACT 701

RESULT 13
US-08-474-379C-17
; Sequence 17, Application US/08474379C
; Patent No. 5977305
; GENERAL INFORMATION:
; APPLICANT: Wigler, Michael H.
; APPLICANT: Colicelli, John J.
; TITLE OF INVENTION: CLONING BY COMPLEMENTATION AND RELATED
; TITLE OF INVENTION: PROCESSES
; NUMBER OF SEQUENCES: 88
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive/6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474,379C
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/511,715
; FILING DATE: 20-APR-1990
; PRIOR APPLICATION DATA:

```

: APPLICATION NUMBER: US 08/206,188
: FILING DATE: 01-MAR-1994
: PRIORITY APPLICATION DATA:
: APPLICATION NUMBER: US 07/688,352
: FILING DATE: 19-APR-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: Clough, David W.
: REGISTRATION NUMBER: 36,107
: REFERENCE/DOCKET NUMBER: 27866/32771
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (312) 474-6300
: TELEFAX: (312) 474-0448
: INFORMATION FOR SEQ ID NO: 17:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1299 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: CDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1..1299
: US-08-474-379C-17

Query Match
Best Local Similarity 100.0%; Score 16; DB 2; Length 1299;
Pred. No. 1.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 960 tgaaggaatcttact 975
Db 686 TGAAGGAATCTTACT 701
|||||

RESULT 14
US-09-146-249A-17
: Sequence 17, Application US/09146249A
: Patent No. 6069240
: GENERAL INFORMATION:
: APPLICANT: Migler, Michael H.
: APPLICANT: Colicelli, John J.
: TITLE OF INVENTION: Cloning by Complementation and Related
: TITLE OF INVENTION: Processes
: NUMBER OF SEQUENCES: 85
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
: STREET: 6300 Sears Tower, 233 South Wacker Drive
: CITY: Chicago
: STATE: Illinois
: COUNTRY: United States of America
: ZIP: 60606-6402
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/146,249A
: FILING DATE:
: CLASSIFICATION:
: PRIORITY APPLICATION DATA:
: APPLICATION NUMBER: US 07/511,715
: FILING DATE: 20-APR-1990
: ATTORNEY/AGENT INFORMATION:
: NAME: Clough, David W.
: REGISTRATION NUMBER: 36,107
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 312/474-6300
: TELEFAX: 312-474-0448
: TELEEX: 25-3856
: INFORMATION FOR SEQ ID NO: 17:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1299 base pairs
```

```

: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: CDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1..1299
: US-09-146-249A-17

Query Match
Best Local Similarity 100.0%; Score 16; DB 3; Length 1299;
Pred. No. 1.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 960 tgaaggaatcttact 975
Db 686 TGAAGGAATCTTACT 701
|||||

RESULT 15
US-08-206-188B-17
: Sequence 17, Application US/08206188B
: Patent No. 6100025
: GENERAL INFORMATION:
: APPLICANT: Migler, Michael H.
: APPLICANT: Colicelli, John J.
: TITLE OF INVENTION: Cloning by Complementation and Related
: TITLE OF INVENTION: Processes
: NUMBER OF SEQUENCES: 84
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
: STREET: 6300 Sears Tower, 233 South Wacker Drive
: CITY: Chicago
: STATE: Illinois
: COUNTRY: United States of America
: ZIP: 60606-6402
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/206,188B
: FILING DATE: 01-MAR-1994
: CLASSIFICATION: 435
: PRIORITY APPLICATION DATA:
: APPLICATION NUMBER: US 07/511,715
: FILING DATE: 20-APR-1990
: ATTORNEY/AGENT INFORMATION:
: NAME: Clough, David W.
: REGISTRATION NUMBER: 36107
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 312/474-6300
: TELEFAX: 312-474-0448
: TELEEX: 25-3856
: INFORMATION FOR SEQ ID NO: 17:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1299 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: CDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1..1299
: US-08-206-188B-17

Query Match
Best Local Similarity 100.0%; Score 16; DB 3; Length 1299;
Pred. No. 1.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 960 tgaaggaatcttact 975
```

Wed May 22, 09:23:35 2002

us-09-693-205-7_copy_5300_6700.oli.rni

Page 7

|||||||
Db 686 TGAAGGAATCTTACT 701

Search completed: May 22, 2002, 08:31:08
Job time: 6993 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 22, 2002, 05:31:27 ; Search time 3619.39 Seconds
(without alignments)
2211.338 Million cell updates/sec

Title: US-09-693-205-7_COPY_12200_12792

Perfect score: 593
Sequence: 1 acatctatgttaccagct.....acctccacatttattgctt 593

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 674847542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_estbm:*
3: em_estlin:*
4: em_estlov:*
5: em_estlov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_estl:*
10: gb_estl2:*
11: gb_hic:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	581	98.0	828	10	BM470780 AGENCOURT
2	565.2	95.3	660	10	BE890125
3	539.6	91.0	632	9	AM968633
4	530.6	89.5	739	9	AL567149
5	456	76.9	543	9	AI932370
6	440	74.2	536	10	BF438152
7	435.4	73.4	536	10	BE856736
8	423	71.3	500	9	AI803488
9	410	69.1	497	9	AM087745
10	374	63.1	469	9	AI217518
11	368	62.1	368	9	AA683013
12	347.6	58.6	483	10	N48291
13	341	57.5	422	9	AI078834
14	330	55.6	410	9	AA954825
15	327.4	55.2	418	10	N46342
16	323	54.5	404	9	AA809783
17	322.4	54.4	416	9	AI499896

18	310	52.3	528	9	AA417817	AA417817 zv04h08.r
19	275.2	46.4	407	9	AA481507	AA481507 aa34c04.s
20	273.6	46.1	441	10	N59442	N59442 yz30b12.s1
21	270.8	45.7	1016	10	BE896315	BE896315 601439161
22	267.8	45.2	349	9	AI561086	AI561086 tq26b04.x
23	252	42.5	351	9	AM262498	AM262498 xq85c06.x
24	240	40.5	456	9	AA417676	AA417676 zv04d08.r
25	204.4	34.5	727	10	BC619034	BC619034 602616564
26	190.2	32.1	1079	10	BM476997	BM476997 AGENCOURT
27	182	30.7	287	10	TI1045	TI1045 NIB26 Norm
28	165.2	27.9	840	10	BF693898	BF693898 602082463
29	157.2	26.5	789	10	BE889418	BE889418 601512535
30	142.4	24.0	302	9	AI865982	AI865982 wX88903.x
31	138.4	23.3	566	10	BM119091	BM119091 L0920E06-
32	138.4	23.3	711	10	BM119546	BM119546 L0926H02-
33	137.6	23.2	611	10	BM247244	BM247244 K0749B07-
34	134.4	22.7	611	9	AM555362	AM555362 L0254H06-
35	134.4	22.7	711	10	BG075163	BG075163 H3144A04-
36	132.8	22.4	543	9	AM555384	AM555384 L0255B07-
37	130.4	22.0	428	10	BE862199	BE862199 UI-M-BHO-
38	125.8	21.2	203	9	AA228047	AA228047 zt58a07.r
39	125.6	21.2	530	10	BM118416	BM118416 L0910F10-
40	119.2	20.1	589	9	AM524447	AM524447 UI-R-B00-
41	106.2	17.9	677	10	BF576822	BF576822 602132834
42	93.8	15.8	460	9	AI506773	AI506773 vM58g10.x
43	85	14.3	402	9	AV666716	AV666716 AV666716
44	83.2	14.0	464	10	BF389741	BF389741 UI-R-BS2-
45	81.8	13.8	529	9	AA124736	AA124736 m982f11.r

ALIGNMENTS

RESULT 1
BM470780 828 bp mRNA linear EST 05-FEB-2002
LOCUS AGENCOURT_6463084 NIH_MGC_71 Homo sapiens cDNA IMAGE:5533575
DEFINITION 5' mRNA sequence.
ACCESSION BM470780
VERSION BM470780.1 GI:18519822
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE NIH-MGC http://mgc.nci.nih.gov/
1 (bases 1 to 828)
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapds-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: L1AM12218 row: k column: 16
High quality sequence stop: 680.
Location/Qualifiers
1. 828
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5533575"
/clone_lid="NIH_MGC_71"
/tissue_type="leptomysarcoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2.1 kb."
262 a 107 c 138 g 321 t

FEATURES

source
BASE COUNT
ORIGIN

Query Match 98.0%; Score 581; DB 10; Length 828;
 Best Local Similarity 99.7%; Pred No. 8.9e-99;
 Matches 593; Conservative 0; Mismatches 0; Indels 2; Gaps 1;

```

QY 1 acatctatgtttacagagcttcctgtttgatgaagatagaacggaactcaaatggt 60
    |||
DB 143 ACATCTTATAGTTTACAGAGCTTCCTGTTTGATGAAGATAGCAACGGAACCTCAAAATGGT 202
QY 61 ggcagttcttatcaacagttgttagtattgtttcttgcgaacagcttgcgaagaacatt 120
    |||
DB 203 GGCAGTTCCTTATACAGAGTTGATGATGTTTCTGGAACCTGCTCCAGCAACACATT 262
QY 121 tattaactgttagaacaactgtcttattgtttgtgttacctattttccacaatgttata 180
    |||
DB 263 TATTAACTGTTAGAACACTTGTCTTATGTTTGTGTGTACATATTTTCCACAAATGTTATA 322
QY 181 attatatagttgttgaacagatgcaactcttctgttcttaaaagtgctgcaatt-- 238
    |||
DB 323 ATTATATAGTGTGTTGAACAGATGCAATCTTTGTTGTCTAAAGTCTGCTGACATTAA 382
QY 239 aaaaaaaaaaacacttcttcttcaataatgagctgtgagtgagttttttaaacttaaaa 298
    |||
DB 383 AAAAAAAAAAACACTTCTTCTTCAATATGCGCATGATGAGTGTGAGCTTTTAACTTAAAA 442
QY 299 aatacaaaaattgttaaatcatctgttatactagtagtataatataatcagcttattat 358
    |||
DB 443 ACATCAAAAATTTTAAATCATGTTGTTATCTAGTAGTTTAAATTAATATGCGCTTATATT 502
QY 359 tccccatgaatgatacgaactgacatttaattcaatgtttgtctgcacatgcttcttact 418
    |||
DB 503 TCCCCATGATGATGATGCAAGCTGACATTTTAAATCATGTTTGTCTGCGCATCTTCTTACT 562
QY 419 ttaacatattcttcttgcgaagaatgataaagtgataataattagttatataagttact 478
    |||
DB 563 TTACATATTTCTTTTGCAGAAATGTAAGTAAGTAAATATTAATTAATTAAGTGTACT 622
QY 479 ggcgttaaatgatacgaactgataacttaatacgaatgaagcttacaagaacatgttgaac 538
    |||
DB 623 GGCCTGAATGATGATGATTAATATCTTATGCAATTTAAGGCTTACAGAACATGTTGAAC 682
QY 539 ttttttactttatttgggaataaagaatgtttgcacccctcacatttattgctt 593
    |||
DB 683 TTTTCTTACTTTTATGGAATGAAGATGTTGACCTCCACATTTTATTGCTT 737

RESULT 2
BE890125 660 bp mRNA linear EST 20-OCT-2000
LOCUS 601513104F1 NIH_MGC_71 Homo sapiens cDNA IMAGE:3914521 5',
DEFINITION mRNA sequence.
ACCESSION BE890125
VERSION BE890125.1 GI:10348134
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 660)
NIH-MGC http://mgi.mc.nhl.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapds-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Prepared by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
plate: L1AM9736 row: 9 column: 02

```

FEATURES
 source High quality sequence stop: 628.
 Location/Qualifiers
 1..660

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3914521"
 /clone_1ib="NIH-MGC_71"
 /tissue_type="Telomerase"
 /lab_host="PH10B (phage-resistant)"
 /note="Organ: uterus; Vector: pCMV-Sport6; Site 1: NotI;
 Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 2.1 kb."

BASE COUNT 196 a 95 c 107 g 262 t
 ORIGIN

Query Match 95.3%; Score 565.2; DB 10; Length 660;
 Best Local Similarity 99.2%; Pred. No. 8.1e-96;
 Matches 589; Conservative 0; Mismatches 3; Indels 2; Gaps 2;

```

QY 1 acatctatgtttacagagcttcctgtttgatgaagatagaacggaactcaaatggt 60
    |||
DB 41 ACATCTTATAGTTTACAGAGCTTCCTGTTTGATGAAGATAGCAACGGAACCTCAAAATGGT 100
QY 61 ggcagttcttatcaacagttgttagtattgtttcttgcgaacagcttgcgaagaacatt 120
    |||
DB 101 GGCAGTTCCTTATACAGAGTTGATGATGTTTCTGGAACCTGCTCCAGCAACACATT 160
QY 121 tattaactgttagaacaactgtcttattgtttgtgttacctattttccacaatgttata 180
    |||
DB 161 TATTAACTGTTAAACACTTGTCTTAAATGTTGTTGTGTACATATTTTCCACAAATGTTATA 220
QY 181 attatatagttgttgaacagatgcaactcttctgttcttaaaagtgctgcaatt-a 239
    |||
DB 221 ATTATATAGTGTGTTGAACAGATGCAATCTTTGTTGTCTAAAGTCTGCTGACATTAA 280
QY 240 aaaaaaaaaaacacttcttcttcaataatgcatgcatgtgagtttctttaaacttaaaa 299
    |||
DB 281 AAAAAAAAAAACACTTCTTCTTCAATATGCGCATGATGAGTGTGAGTGTTTTAACTTAAAA 340
QY 300 catcaaaaattgttaaatcatctgttcttctctatctgtattataatcagcttattat 359
    |||
DB 341 CATCAAAAATTTGTTAAATCATGTTGTTATGTTTAAATTAATTAATTAATTAATTT 400
QY 360 ccccatgaatgatacgaactgacatttaattcaatgtttgtctgcacatgcttcttact 419
    |||
DB 401 CCCCATGAATGATGATGATTAATATCTTATGCAATTTAAGGCTTACAGAACATGTTGAAC 460
QY 420 taacataattcttcttgcgaagaatgataaagtgataatgaattatataaagtactg 479
    |||
DB 461 TAACATATTTCTTTGCAAGATGATGAAGTAAATTAATTAATTAATTAATTAATTAAT 520
QY 480 gctgtaaatgatacgaacttatacttataatgaatgaagcttacaagaatgttgaact 539
    |||
DB 521 GCTGTAATGATGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 580
QY 540 ttttttactttatttgggaataaagaatgtttgcacccctcacatttattgctt 593
    |||
DB 581 TTTTCTTACTTTTATGCGAAT-AGGAATGTTTCCACTCCACATTTATTGCTT 633

RESULT 3
AW968633 632 bp mRNA linear EST 01-JUN-2000
LOCUS AW968633
DEFINITION EST180709 MAGE resequences, MAGE Homo sapiens cDNA, mRNA sequence.
ACCESSION AW968633
VERSION AW968633.1 GI:8158474
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 632)

```



```

Db      119 AACATTTCTTTGCAAGATGTAAGTAATGATTAGTTATATAGTCTACTCG 60
Oy      481 ctgtaaatgatgctaataatacttatacttgaatgaaggtctacagaacatgttgaact 539
Db      59  CTTGAATGATGCTAATATACCTTTATGCAATTAAGGCTTACAGAACATGTTGACT 1

RESULT  5
LOCUS   AI932370      543 bp      mRNA      linear      EST 17-DEC-1999
DEFINITION wd27ell.x1 Soares_NFL_T.GBC.S1 Homo sapiens cDNA clone
ACCESSION AI932370
VERSION   AI932370.1 GI:5671107
KEYWORDS EST.
SOURCE   human.
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 543)
AUTHORS  NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
          Tumor Gene Index
JOURNAL  Unpublished (1997)
COMMENT  Contact: Robert Strausberg, Ph.D.
          Email: cgapbs-r@mail.nih.gov
          This clone is available royalty-free through LNLN; contact the
          IMAGE Consortium (info@image.llnl.gov) for further information.
          Insert length: 982 Std Error: 0.00
          Seq primer: -400P from Gibco
          High quality sequence stop: 455.
          Location/Qualifiers
            1..543
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /clone="IMAGE:232938"
              /clone_1lb="Soares_NFL_T.GBC.S1"
              /lab_host="DH10B"
              /note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with
              a modified polylinker; Site.1: Not I; Site.2: Eco RI;
              Equal amounts of plasmid DNA from three normalized
              libraries (fetal lung NBHL19W, testis NHT, and B-cell
              NCI-CGAP-GCB1) were mixed, and ss circles were made in
              vitro. Following HAP purification, this DNA was used as
              tracer in a subtractive hybridization reaction. The driver
              was PCR-amplified cDNAs from pools of 5,000 clones made
              from the same 3 libraries. The pools consisted of
              I.M.A.G.E. clones 297480-302087, 682632-687239,
              726408-728711, and 723096-731399. Subtraction by Bento
              Soares and M. Fatima Bonaldo."
BASE COUNT  215 a      86 c      70 g      167 t      5 others
ORIGIN

Query Match      76.9%; Score 456; DB 9; Length 543;
Best Local Similarity 99.8%; Pred. No. 1,8e-75;
Matches 467; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Oy      126 acgtttaagaacacttgcttaattgtgtgtgtacataatttcacaaatttaattta 185
Db      543 ACTGTTAGAACCTTCCTTATGTTGTGTGTCATATTTTCCACAAATGTTATTAATTTA 484
Oy      186 tatagtgtgttaagaacagatgcaactcttgtgtcttaagaagtgtcgcagttaaaaaa 245
Db      483 TATAGTGTGTTGACAGAGATGCAATCTTTGTGTCTTAAGGTCGCGCATTTAAAAAAA 424
Oy      246 aaacaaccttctctcaatatagcatgtagtgagtttttaacttaaaaaacatcaa 305
Db      423 AAACAACCTTTCTTCATATGCAATGATGAGAGTTTAACTTTAAAAACATCAAA 364
Oy      306 aaatgttaaatcatgtgtatctagtagttataataatataatggtctatattcccat 365
Db      303 AAATGTTTAAATCATGTTGTATGTAGTAGTTTATTAATATATCGGCTTATATTTCCCAT 304

```

```

Oy      366 gaatgatcagaacatgacatlttaattcatgttgcgcgcagtgcttcttaacttaacat 425
Db      303 GAATGATCAGAACATGACATTTTAATTCATGTTGTCTCGCCATGCTTTCTTAATCAAT 244
Oy      426 attccttttcagagaatgtaaaaggtaatgataatagttatataaagtgtacgtgcgtga 485
Db      243 ATTTCTTTTGCGAATGTAAAAAGGTAATGATTAATGTTATTAATAGTACTGCGTGTGA 184
Oy      486 aatgatcctaataacttcttgcgaatgaaggtcttcacagacatgttgaaccttttt 545
Db      183 AATGATGCTTAATTAATCTTTATGCAATTAAGGCTTACAGAACATGTTGAAC-TTTTTT 125
Oy      546 accttataatggaataaagaatgtttgcacccacacattatgcct 593
Db      124 ACTTTATTTGGGAATFAAGGAATGTTTGACCTCCACATTTATTTGCTT 77

RESULT  6
LOCUS   BF438152/c      536 bp      mRNA      linear      EST 30-MAR-2001
DEFINITION 7667f12.x1 NCI-CGAP_Lu24 Homo sapiens cDNA clone IMAGE:3703462 3',
ACCESSION BF438152
VERSION   BF438152.1 GI:11450669
KEYWORDS EST.
SOURCE   human.
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 536)
AUTHORS  NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
          Tumor Gene Index
JOURNAL  Unpublished (1997)
COMMENT  Contact: Robert Strausberg, Ph.D.
          Email: cgapbs-r@mail.nih.gov
          Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
          Emmert-Buck, M.D., Ph.D.
          CDNA Library Preparation: M. Bento Soares, Ph.D.
          CDNA Library Arrayed by: Greg Lennon, Ph.D.
          DNA Sequencing by: Washington University Genome Sequencing Center
          Clone distribution: NCI-CGAP clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL, send email to:
          info@image.llnl.gov
          Seq primer: -400P from Gibco
          High quality sequence stop: 481.
          Location/Qualifiers
            1..536
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /clone="IMAGE:3703462"
              /clone_1lb="NCI-CGAP_Lu24"
              /tissue_type="carcinoid"
              /lab_host="DH10B"
              /note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a
              modified polylinker; Plasmid DNA from the normalized
              library NCI CGAP_Lu24 was prepared, and ss circles were
              made in vitro. Following HAP purification, this DNA was
              used as tracer in a subtractive hybridization reaction.
              The driver was PCR-amplified cDNAs from a pool of 5,000
              clones made from the same library (clonoids
              141920-1417991 and 1520904-1522439). Subtraction by Bento
              Soares and M. Fatima Bonaldo."
BASE COUNT  212 a      85 c      67 g      172 t
ORIGIN

Query Match      74.2%; Score 440; DB 10; Length 536;
Best Local Similarity 99.8%; Pred. No. 1,7e-72;
Matches 451; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Oy      143 ttatgtttgtgtlacataatttcacaaatgttataatattatagtggtgtgaaca 202

```

```
|||||
Db 536 TTATGTTGGTGTACATATTTTCCAAATGTTATATATATAGTGTGGTGAACA 477
Qy 203 ggaatgaacattcttctgtcttaagaagtcgcagct-aaaaaaaacaaccttctt 261
Db 476 GGATGCATCTTTTGTGTCTAAAGTGTGCAGTTAAAAAACAACCTTTCTT 417
Qy 262 caataagcattgtagagagtttttaactttaaaacatcaaaatctgttaaatcat 321
Db 416 CAATATGCAATGTAAGTGAAGTTTAACTTTAAAAACATCAAAAATGTTAAATCAT 357
Qy 322 tctgttaactagtagtataataatcagctataatctcccatgaatgaacagactga 381
Db 356 TGGTTATTCAGTAGCTTTATATATATGCGCTATATTTCCCATGAATGATCAGACGA 297
Qy 382 caattaatcattgtctgcgcagctctcttaactttaacataattctttgcagaat 441
Db 296 CATTATATTCATGTTTGTCTCGCATGCTCTTACTTTAACATATTTCTTTGCAGAA 237
Qy 442 gtaaaagtaatgtaattgattatataagtgactgcgcgtgaatgaatgaataata 501
Db 236 GTAAAGGTATGATATATTTGTTATATTAAGTACTGCTGTAAATGATCTAAATTA 177
Qy 502 cttaatacaatlaaggcttaacagaacatgtgaaccttttcaactttatgtggaata 561
Db 176 CTTATTCATTAAGGCTTACAGACATGTTGAACCTTTTATCTTTATTTGGAATA 117
Qy 562 aggaatgttcgacctccacatttatgctt 593
Db 116 AGGAATGTTGCACCTCCACATTTTATGCTT 85

RESULT 7
BE856736/c 536 bp mRNA linear EST 29-SEP-2000
LOCUS 7f68a06.x1 Soares_NSF_F8_9W_OT_PA_P_S1 Homo sapiens cDNA clone
DEFINITION IMAGE:3299794 3', mRNA sequence.
ACCESSION BE856736
VERSION BE856736.1 GI:10370063
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 536)
AUTHORS NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLES National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
JOURNAL Tumor Gene Index
COMMENT Unpublished (1997)
CONTACT: Robert Strausberg, Ph.D.
Email: cgaaps-f@mail.nih.gov
This clone is available royalty-free through LINT; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -400P from Glibco
High quality sequence stop: 448.
Location/Qualifiers
1..536
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3299794"
/clone_id="Soares_NSF_F8_9W_OT_PA_P_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pVT73D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Equal amounts of plasmid DNA from five normalized
libraries were mixed, and ss circles were made in vitro.
Following HAP purification, this DNA was used as tracer in
a subtractive hybridization reaction. The driver was from
PCR-amplified cDNAs from pools of 5,000 clones made from
the same 5 libraries. The pools consisted of the following
libraries and cloneIDs: Soares NBHSF pool 1:
309384-310919, 323208-325895 Soares NB2HP pool 1:
145032-147335, 147720-148103, 148872-149255, 15002 -
```

```
150407, 151176-152327 Soares NB2HP-9W pool 1:
758280-760583, 772104-774407 Soares NBHPA pool 1:
304776-306311, 320136-322823, 326280-326663 Soares NBHOT
pool 1: 723720-726407, 739080-740999 Subtraction by Bento
Soares and M. Fatima Bonaldo."
BASE COUNT 209 a 85 c 68 g 174 t
ORIGIN
Query Match 73.4%; Score 435.4; DB 10; Length 536;
Best Local Similarity 99.6%; Pred. No. 1.2e-71;
Matches 447; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
Qy 146 atgttggtgtacattcttcacaaatgtaataattatataagtggtgaacagga 205
Db 536 ATGTGTGTGTACATATTTTCCAAATGTTATATATATAGTGTGGTGAACAGA 477
Qy 206 tgaactcttctgtcttaagaagtcgcagct-aaaaaaaacaaccttcttca 264
Db 476 TGCATCTTTTGTGTCTAAAGTGTGCAGTTAAAAAACAACCTTTCTTCA 417
Qy 265 tatgcatagtagtgagtttttaactttaaaacatcaaaatctgttaaatcattgt 324
Db 416 TATGCATGTAAGTGAAGTTTAACTTTAAAAACATCAAAAATGTTAAATCATG 357
Qy 325 gttaactagtagtataataatcagctataatctcccatgaatgaatgaacatgacat 384
Db 356 GTATCTGTAGTATATATATATGCGCTTATATTTTCCCATGAATGATGAGTACAT 297
Qy 385 ttaactagtagtgcgcgcagctcttcaactttaacataattctttgagaatga 444
Db 296 TTAATTCATGTTTGTCTCGCATGCTCTTCTTAAATATTTCTTTTGGAAATGA 237
Qy 445 aaaggtaatgaatlaattatataaagtgactgcgtgtaaaagatgcttaataact 504
Db 236 AAAGTATGATTAATTAATTAATTAAGTACTGCTGCTAAATGATGCTTAATTA 177
Qy 505 tatgcaatlaaggcttaacagaacatgtgaaccttttcaacttttttgggaataag 564
Db 176 TATGCATTAAGGCTTACAGACATGTTGAACCTTTTATCTTTATTTGGAATAAG 117
Qy 565 aatgttcgacctccacatttatgctt 593
Db 116 AATGTTGCACCTCCACATTTTATGCTT 88

RESULT 8
A1803488/c 500 bp mRNA linear EST 13-DEC-1999
LOCUS tc17g02.x1 Soares_NHMPu_S1 Homo sapiens cDNA clone IMAGE:2064146
DEFINITION 3', mRNA sequence.
ACCESSION A1803488
VERSION A1803488.1 GI:5368882
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 500)
AUTHORS NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLES National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
JOURNAL Tumor Gene Index
COMMENT Unpublished (1997)
CONTACT: Robert Strausberg, Ph.D.
Email: cgaaps-f@mail.nih.gov
This clone is available royalty-free through LINT; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1089 Std Error: 0.00
Seq primer: -400P from Glibco
High quality sequence stop: 447.
Location/Qualifiers
1..500
/organism="Homo sapiens"
```

```

/db_xref="taxon:9606"
/clone.lib="Soares.NhHMPU.S1"
/tissue_type="pooled human melanocyte, fetal heart, and
pregnant uterus"
/lab_host="DH10B"
/Note="Organ: mixed (see below); Vector: pT73D-Pac
(pharmacia) with a modified polylinker; Site.1: Not I;
Site.2: Eco RI; Equal amounts of plasmid DNA from three
normalized libraries (melanocyte 2NH19M, pregnant uterus
NHHPU, and fetal heart NBH19M) were mixed, and ss circles
were made in vitro. Following HAP purification, this DNA
was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from pools of
5,000 clones made from the same 3 libraries. The pools
consisted of I.M.A.G.E. clones 260232-265223,
340488-345479, and 484488-489479."
BASE COUNT      196 a      80 c      62 g      162 t
ORIGIN

```

```

Query Match      71.3%; Score 423; DB 9; Length 500;
Best Local Similarity 100.0%; Pred. No. 2.5e-69;
Matches 423; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 171 aatgtataattatagtggtgtgacagatcattctgtgtctaaagtg 230
|||||
DB 500 AATGTTATATATATATAGTGTGTTGAACAGATGCAATCTTTGTGCTAAAGTG 441
|||||
QY 231 ctgcagttlaaaaaaacacattcttccaatgcatgcatgagatttttaa 290
|||||
DB 440 CTCGAGTTAAAAAACAACCTTTCTTCAATATGATGATGAGAGTTTAA 381
|||||
QY 291 cttlaaaacataaaatgtaaaatcattggtatctagtaagtttaatacgg 350
|||||
DB 380 CTTTAAAAACATCAAAATGTTAAATCATGTGTATGATGATTAATATCGG 321
|||||
QY 351 ctatacttcccatgaatgacagactacatttaattcattgtctgcagatgct 410
|||||
DB 320 CTTATATTTCCCATGATGATGATGATGATGATGATGATGATGATGATG 261
|||||
QY 411 tcttacccttaacattcttctgcagaatgtaaaagtaagtaataagttata 470
|||||
DB 260 TCTTTACTTTAACATATTTCTTTTGCAGATGTAAGGTAATGATTAATTA 201
|||||
QY 471 agtgaactgctgaatgctgaatgctgaatgctgaatgctgaatgctgaatg 530
|||||
DB 200 AGGTACTGCTGTAATGATGATGATGATGATGATGATGATGATGATGATG 141
|||||
QY 531 gtgaacacttcttcttcttcttcttcttcttcttcttcttcttcttcttct 590
|||||
DB 140 GTTGAACCTTTTACTTTATTTATTTATTTATTTATTTATTTATTTAT 81
|||||
QY 591 ctt 593
|||||
DB 80 CTT 78

```

```

RESULT 9
AM087745/c 497 bp mRNA linear EST 15-OCT-1999
LOCUS      xB6F08.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
DEFINITION IMAGE:2581479 3', mRNA sequence.
ACCESSION AM087745
VERSION    AM087745.1 GI:6043550
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens

```

```

REFERENCE 1 (bases 1 to 497)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

```

```

JOURNAL    Tumor Gene Index,
            Unpublished (1997)
COMMENT    Contact: Robert Strausberg, Ph.D.
            Email: cgaps-remail.nih.gov
            This clone is available royalty-free through LNL; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -400P from Glibco
            High quality sequence stop: 455.
            Location/Qualifiers

```

```

FEATURES   1..497
            source

```

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone.lib="Soares.NhHMPU.S1"
/lab_host="DH10B"
/Note="Organ: pooled; Vector: pT73D-Pac (pharmacia) with
a modified polylinker; Site.1: Not I; Site.2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NBH19M, testis NHT, and B-cell
NCI-CGAP GCB1) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Benito
Soares and M. Fatima Bonaldo."
BASE COUNT      195 a      80 c      63 g      159 t
ORIGIN

```

```

Query Match      69.1%; Score 410; DB 9; Length 497;
Best Local Similarity 99.8%; Pred. No. 6.6e-67;
Matches 421; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

```

```

QY 172 aatgtataattatagtggtgtgacagatcattctgtgtctaaagtg 231
|||||
DB 497 AATGTTATATATATATAGTGTGTTGAACAGATGCAATCTTTGTGCTAAAGTG 438
|||||
QY 232 tgcagttlaaaaaaacacattcttccaatgcatgcatgagatttttaa 291
|||||
DB 437 TCGAGTTAAAAAACAACCTTTCTTCAATATGATGATGATGATGATGATG 378
|||||
QY 292 cttlaaaacataaaatgtaaaatcattggtatctagtaagtttaatacgg 351
|||||
DB 377 TTTAAAAACATCAAAATGTTAAATCATGTGTATGATGATGATGATGATG 318
|||||
QY 352 ttatacttcccatgaatgacagactacatttaattcattgtctgcagatgct 411
|||||
DB 317 TTATATTTCCCATGATGATGATGATGATGATGATGATGATGATGATG 258
|||||
QY 412 cttaacttcaacattcttctgcagaatgtaaaagtaagtaataagttata 471
|||||
DB 257 CTTTACTTTTAAATATTTCTTTTGCAGATGTAAGGTAATGATTAATTA 198
|||||
QY 472 gtgtactgctgaatgctgaatgctgaatgctgaatgctgaatgctgaatg 531
|||||
DB 197 GTGTACTGCTGTAATGATGATGATGATGATGATGATGATGATGATGATG 138
|||||
QY 532 ttgaacacttcttcttcttcttcttcttcttcttcttcttcttcttct 591
|||||
DB 137 TTGAACCTTTTACTTTATTTATTTATTTATTTATTTATTTATTTAT 79
|||||
QY 592 tt 593
|||||
DB 78 TT 77

```

```

RESULT 10
AI217518/c 469 bp mRNA linear EST 17-MAR-1999
LOCUS      qh20g08.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
DEFINITION IMAGE:1845278 3', mRNA sequence.

```

ACCESSION AI217518
VERSION AI217518.1 GI:3797333
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
AUTHORS 1 (bases 1 to 469)
TITLE NCI-CCAP http://www.ncbi.nlm.nih.gov/nciccap.
JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
COMMENT Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaaps-remail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1085 Std Error: 0.00
Seq primer: -400p from Glibco
High quality sequence stop: 456.
Location/Qualifiers
1..469
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1845278"
/clone_lib="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NbH19W, testis NHT, and B-cell
NCI-CCAP GCB1) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo."
BASE COUNT 184 a 71 c 63 g 151 t
ORIGIN
Query Match 63.1%; Score 374; DB 9; Length 469;
Best Local Similarity 99.7%; Pred. No. 3.3e-60;
Matches 385; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
QY 209 aatccttctgtctaaagtgctgcagtttaaaaaa-caaccttcttccaat 267
|||||
DB 469 AATCTTTGTTGCTTAAGTGCGAGTAAACCAACACCTTTCTTCAATAT 410
QY 268 ggcctgtagtgaggtcttcttaacttaaaacatcaaaatgtttaaattgt 327
|||||
DB 409 GGCATGTAGTGGAGTTTAACTTAAACATCAAAATTTGTTAAATCATGTGTT 350
QY 328 atcgaagagtttaaatatcaggtcttatctcccatcgaatgctcagaatgacatt 387
|||||
DB 349 ATCTAGAGATTATTAATATGCGCTTATTTCCCATGAATGATCAAGACGACATTAT 290
QY 388 attcaatgttgcctcgcacatgctcttacttaacatattcttcttcagaatgtaa 447
|||||
DB 289 ATTCATGTTTGTCGCGCATGCTTCTTACTTAAACATATTTCTTTTCAGAAATTA 230
QY 448 ggtataataatagttataatagtgactgctgtaaatgataatataattat 507
|||||
DB 229 GGTAAATATATATGTTTATTAAGTACTGCTGTAAGATGCTAAATATATCTTAT 170
QY 508 gcaatlaagagctcagaagaatgttgaacttttacttatttgggaataagat 567
|||||
DB 169 GCAATTAAGGCTTACAGAACATGTTGAACCTTTTCTTATTTATGGAATTAAGAT 110
QY 568 gttgcacctccacatttattgtct 593
|||||
DB 109 GTTGCACCTCCACATTTTATGCTT 84

RESULT 11
AA683013/c 368 bp mRNA linear EST 15-DEC-1997
LOCUS
DEFINITION aeg1b08.s1 Stratiogene schizo brain S11 Homo sapiens cDNA clone
IMAGE:970551 3', mRNA sequence.
ACCESSION AA683013
VERSION AA683013.1 GI:2668904
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
AUTHORS 1 (bases 1 to 368)
Hillier, L., Allen, M., Bowles, L., Dubuque, T., Giesel, G., Jost, S.,
Kizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin
J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B.,
White, Y., Wylie, T., Waterston, R. and Wilson, R.
TITLE Washu-NCI human EST Project
JOURNAL Unpublished (1997)
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Possible reversed clone: poly not found
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 358.
Location/Qualifiers
1..368
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:970551"
/clone_lib="Stratiogene schizo brain S11"
/sex="male"
/tissue_type="schizophrenic brain S-11 frontal lobe"
/dev_stage="34 years old"
/lab_host="SOLR (kanamycin resistant)"
/note="Vector: Bluescript SK-; Site_1: EcoRI. Library
constructed from S-11 frontal lobe, male, 34 years old,
50% caucasian, 50% Aleutian. Schizophrenic suicide.
Random primed into EcoRI site of ZAP II Vector. Mass
excised. Custom library. Avg insert length 1.4kb.
Material obtained by Johnston N., Torrey, E.F., Yolken R.,
and the Stanley Neuropathology Consortium - Analysis of
RNAs from the Brains of Individuals with Psychiatric
Diseases (unpublished) Stanley Neurovirology Laboratory,
Johns Hopkins School of Medicine, Baltimore MD."
BASE COUNT 140 a 61 c 52 g 115 t
ORIGIN
Query Match 62.1%; Score 368; DB 9; Length 368;
Best Local Similarity 100.0%; Pred. No. 4.6e-59;
Matches 368; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 35 gatacaacggaactcaaatggtgagcttctataccagttgttaagtgttc 94
|||||
DB 368 GATACCAACGGAAACCTCAAAATGTGCGAGTTTATTAACAGTTGTAAGTATGTTTC 309
QY 95 tgaagcgtctgccaagaacaattatcaactgttgaacaactgcttctgttg 154
|||||
DB 308 TCGAAGTCGCTTGCCAGACACATTTATTAACGTGAACACTGCTTATGTTGCG 249
QY 155 tgtacatatcttccacaatgltataattatatagtgtgttggaacagatgcact 214
|||||
DB 248 TGTACATATTTTCCCAAAATGTTATTAATTAATAGTGTGTTGAACAGATGCAATCTT 189
QY 215 ttgtgtctaaagtgctgcagtttaaaaaaacaacttcttccaatagtcagt 274

DB	168	TTGTGCTGAAGGCGCTGCACTTTAAAAAACAACCTTTCTTTCAATGTGCATCT	129
OY	275	agtggagttttttaactttaaaacatcaaaaattgttaaaatcatgtgttattcagtc	334
Db	128	AGTGAGGTTTTTTTTTAACCTTTAAAAACATCAAAAAATTGTTAAATCATGTGTATCTAGT	69
OY	335	agtttaaatatctgcgtattatctcccatgaatgatcaagaacgtacatttaactagc	394
Db	68	AGTTATTAATTATTCGGCTTATTTATTTCCCATGAAATGATCAGACATGACATTAATTCATAG	9
OY	395	tttgctc 402	
Db	8	TTTGCTC 1	
RESULT	12		
LOCUS	N48291	483 bp	EST 14-FEB-1996
DEFINITION	yy77c12.s1 Soares_multiple_sclerosis.2NBMHSP Homo sapiens cDNA		
ACCESSION	N48291		
VERSION	N48291.1	GI:1189457	
KEYWORDS	EST.		
ORGANISM	Homo sapiens		
SOURCE	human.		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (bases 1 to 483) Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Maizra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevisakis,E., Waterston,R., Williamson,A., Woldmann,P. and Wilson,R. The Mashu-Merck EST Project Unpublished (1995)		
TITLE	Contract: Wilson RK		
JOURNAL	Washington University School of Medicine		
COMMENT	4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Seq primer: ml3 -40 forward High quality sequence stop: 205. Location/Qualifiers 1. .483 /organism="Homo sapiens" /db_xref="GDB:3887950" /db_xref="taxon:9606" /clone="IMAGE:279574" /clone_id="Soares_multiple_sclerosis_2NBMHSP" /sex="male" /tissue_type="multiple sclerosis lesions" /dev_stage="Age 46" /lab_host="DH10B (ampicillin resistant)" /note="Vector: pT73D (Pharmacia) with a modified polylinker V_type: phagemid; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTCACCAATCTGAAGGAGCGCGCCGCACTTTTCTTTTCTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M.Fatima Bonaldo. RNA from 4 multiple sclerosis lesions from one patient was kindly provided by Dr. Kevin G. Becker (NINDS/NIH)."		
BASE COUNT	177 a	80 c	71 g 147 t 8 others

Query Match	Best Local Match	Matches 380:	Similarity	Score	347.6:	DB 10:	Length 483:
				93.6%:	Posed. No. 2.7e-55:		
				0:	Mismatches 22:	Indels	Gaps 2:
OY	192	gtgtgtgaacagaaatgcacatcttctgtctctaaaggtcgtcagttacaaaaaaca	251				
Db	442	GGGNTGGACNCGATGCCATCTTTGTGTGGNCCAAAGCGTCCAGTTAAAAAACC	423				
OY	252	cc---tttcttccaatbgtg-catgtgtgtgagttttttaaacttaaaaacatcaaa	307				
Db	422	ACCCCTTTCCCTTCAATGTGNCACAGTGTGGAGCTTTTAACTTNAACATCAAAA	363				
OY	308	attgttaaatcatgtgttattctagttagttataaatatcgcgttatattcccaaga	367				
Db	362	ATTGTTAAATCAATGTGTATTACTAGNAGTTATATATATATGAGCTTATATTTCC	303				
OY	368	atgacacagaactgcacatttaattcaatgttctctgcacatgcttcttaacttaacat	427				
Db	302	ATGATCAGAACACGACATTTAATTCATGTTTGTGNCGCCATGCTTCTTAAACATAT	243				
OY	428	ttctcttcagaatgtaaagaagtaatgataattatataaagtcatcgtcgtgtaaa	487				
Db	242	NCTCTTTGCAGATGTAAAGATGATATGATTAATGATTATATAGTGTACTGCGCTGT	183				
OY	468	tgatcttaataatactattatgcataatgaaggtctacagaaacatgttgaactttttac	547				
Db	182	TGATCTCTAAATATACCTTATATGCAATTAAGGCGCTTACACAAACATGTGTGAAC	123				
OY	548	ttttcttggaataaggaatgttgcacctcacatttatgctt	593				
Db	122	TTTTATTGGGAATTAAGGAATGTTTGCACCTCCACATTTATTATGCTT	77				
RESULT 13							
LOCUS	A1078834/C						
DEFINITION	A1078834	422 bp	mRNA	linear	EST 10-AUG-1998		
ACCESSION	0246005.x1	Soares_NhMMPu_S1	Homo sapiens	CDNA clone	IMAGE:1678377		
VERSION	A1078834						
KEYWORDS	A1078834.1	GI:3413141					
SOURCE	EST.						
ORGANISM	human.						
	Homo sapiens						
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;						
	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.						
REFERENCE	1 (bases 1 to 422)						
AUTHORS	NCI-CGAP	http://www.ncbi.nlm.nih.gov/ncicgap.					
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),						
JOURNAL	Tumor Gene Index						
COMMENT	Unpublished (1997)						
	Contact: Robert Strausberg, Ph.D.						
	Email: cgapbs-remail.nih.gov						
	This clone is available royalty-free through LNL; contact the						
	IMAGE Consortium (info@image.llnl.gov) for further information.						
	Seq primer: -40ml3 fwd, ET from Amersham						
	High quality sequence stop: 361.						
FEATURES	Location/Qualifiers						
source	1..422						
	/organism="Homo sapiens"						
	/db_xref="taxon:9606"						
	/clone="IMAGE:1678377"						
	/clone_id="Soares_NhMMPu_S1"						
	/tissue_type="Pooled human melanocyte, fetal heart, and						
	pregnant uterus"						
	/lab_host="DH10B"						
	/note="Organ: mixed (see below); Vector: pRT73D-Pac						
	(Pharmacia) with a modified polylinker; Site_1: Not I;						
	Site_2: Eco RI; Equal amounts of plasmid DNA from three						
	normalized libraries (melanocyte 2NbHm, pregnant uterus						
	NbHpu, and fetal heart NbH119w) were mixed, and ss circles						
	were made in vitro. Following HAP purification, this DNA						
	was used as tracer in a subtractive hybridization						
	reaction. The driver was PCR-amplified cDNAs from pools of						


```
/db_xref="taxon:9606"  
/clone_lib="Soares_multiple_sclerosis_2NbHMSp"  
/sex="male"  
/tissue_type="multiple sclerosis lesions"  
/dev_stage="Age 46"  
/lab_host="DH10B (ampicillin resistant)"  
/note="Vector: pT73D (Pharmacia) with a modified  
polylinker V_type: phagemid: Site_1: Not I; Site_2: Eco RI  
; 1st strand cDNA was primed with a Not I - oligo(dT)  
primer [5',  
TGTACCAATCGAAGTGGAGCGCGCATTTTCTTTTCTTTT 3'],  
double-stranded cDNA was size selected, ligated to Eco RI  
adapters (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of a modified pT73 vector  
(Pharmacia). Library went through one round of  
normalization to a Cot = 5. Library constructed by Bento  
Soares and M. Fatima Bonaudo. RNA from 4 multiple sclerosis  
lesions from one patient was kindly provided by Dr. Kevin  
G. Becker (NINDS/NIH)."  
BASE COUNT 167 a 65 c 55 g 130 t 1 others  
ORIGIN
```

```
Query Match 55.2%; Score 327.4; DB 10; Length 418;  
Best Local Similarity 99.1%; Pred. No. 1.6e-51;  
Matches 339; Conservative 0; Mismatches 2; Indels 1; Gaps 1;  
QY 253 cttcttccaatgcatgcatgagagtttcttaacttaaaacacaaatgt 312  
DB 418 CTTTCTTCAATGCGCATGTAGTGGAGCTTTTAACTTMAAACAACAAAATTGP 359  
QY 313 taaatcatgtgtatcctgaagtaataatcagcttatattcccatgaatgat 372  
DB 358 TAAATCATGTGTATCTAGTATTAATTAATGCGCTTATATTTCCCATGAATGAT 299  
QY 373 cagaactgacattcaatcatgtttgctcgcacatgcttcttaacttaacatctt 432  
DB 298 CAGAACTGACATTATATGTTTGTCCGCGCATGCTTACTTAAACATTTCTT 239  
QY 433 ttgcagaatgtaaaaggtaataatgattatataaagtgactgctgtaaatgatg 492  
DB 238 TTGCAGAAATGTAAAGGTATATATATTAATTAAGTACTGCTGTAATGATG 179  
QY 493 ctaaatatacttatgac-aattaaggcttaagaacatgtaaaactttttactttt 551  
DB 178 CTTAAATATACCTTATGCGCAATTAAGGCTTACAGAACATGTGAAACTTTTACTTT 119  
QY 552 attggaataaagaatgttgcacccacacatttatgctt 593  
DB 118 ATTGGGAATTAAGAAATGTTGCACCTCCACATTTTATTCCTT 77
```

Search completed: May 22, 2002, 05:31:36
Job time: 4096 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 22, 2002, 06:53:41 ; Search time 3530.57 Seconds
(without alignments)
3514.853 Million cell updates/sec

Title: US-09-693-205-7_COPY_12200_12792

Perfect score: 593
Sequence: 1 acatctatgtttacagctc.....acctccacatttattgctt 593

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_hlg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_hlg_hum:*
31: em_hlg_inv:*
32: em_hlg_other:*
33: em_hgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result Query
No. Score Match Length DB ID Description

1	593	100.0	4318	9	AB018273	AB018273 Homo sapi
2	593	100.0	12793	6	AX119931	AX119931 Sequence
3	593	100.0	12793	9	AF193556	AF193556 Homo sapi
4	593	100.0	92693	9	AL157766	AL157766 Human DNA
5	593	100.0	99819	2	AC079761	AC079761 Homo sapi
6	426.6	71.9	3289	9	AB056815	AB056815 Macaca fa
7	327.4	55.2	418	11	G36555	G36555 SHGC-53325
8	327.4	10.3	169794	2	AC004688	AC004688 Plasmodu
9	61	10.3	196149	2	AC004709	AC004709 Plasmodu
10	59.6	10.1	4601	3	DMU11584	U11584 Drosophila
11	59.6	10.1	19517	3	DMU37541	U37541 Drosophila
12	57.4	9.7	12029	3	AE001426	AE001426 Plasmodu
13	57.2	9.6	60604	2	AC023466	AC023466 Homo sapi
14	57	9.6	183813	2	AC012204	AC012204 Homo sapi
15	57	9.6	197225	2	AC093835	AC093835 Homo sapi
16	56.8	9.6	110000	2	PFMAL4P1_2	Continuation (3 of
17	56.6	9.5	11691	6	AX347143	AX347143 Sequence
18	56.6	9.5	349980	6	AX344564	AX344564 Sequence
19	55.4	9.3	124635	9	AP000593	AP000593 Homo sapi
20	55.2	9.3	1815	1	AF267220	AF267220 Candidatu
21	55.2	9.3	162075	9	HS127D3	AL021026 Homo sapi
22	55	9.3	1141	6	AX083744	AX083744 Sequence
23	55	9.3	13038	6	AX346176	AX346176 Sequence
24	55	9.3	145598	9	AC008132	AC008132 Homo sapi
25	54.8	9.2	15387	6	AX345086	AX345086 Sequence
26	54.6	9.2	4611	3	FEA132006	AI133006 Plasmodu
27	54.4	9.2	7918	8	AF325123	AF325123 Arabidops
28	54.4	9.2	13574	6	AX346218	AX346218 Sequence
29	54.4	9.2	62268	9	HS1178121	AL109852 Human DNA
30	54.4	9.2	141748	8	AC009323	AC009323 Arabidops
31	54.4	9.2	187048	9	AC024921	AC024921 Homo sapi
32	54	9.1	349980	6	AX344558	AX344558 Sequence
33	54	9.1	349980	6	AX344559	AX344559 Sequence
34	53.8	9.1	863	11	CNS06EVO	AL395628 T7 end of
35	53.8	9.1	189790	2	AC107420	AC107420 Homo sapi
36	53.4	8.9	186625	9	AC006487	AC006487 Homo sapi
37	53	8.9	141469	9	AL139811	AL139811 Human DNA
38	53	8.9	161384	2	AC083801	AC083801 Homo sapi
39	52.8	8.9	17848	6	AX277865	AX277865 Sequence
40	52.8	8.9	17848	6	AX323550	AX323550 Sequence
41	52.8	8.9	17848	6	AX348363	AX348363 Sequence
42	52.8	8.9	19087	6	AX345694	AX345694 Sequence
43	52.8	8.9	77835	9	PFMAL13P2_3	Continuation (4 of
44	52.6	8.9	150754	9	AC023491	AC023491 Homo sapi
45	52.6	8.9	175358	9	AC007981	AC007981 Homo sapi

ALIGNMENTS

RESULT 1
AB018273 4318 bp mRNA linear PRI 16-JUN-1999
LOCUS Homo sapiens mRNA for KIAA0730 protein, partial cds.
DEFINITION AB018273
ACCESSION
VERSION
KEYWORDS
SOURCE
AB018273.1 GI:3882180

Homo sapiens adult male brain cDNA to mRNA, clone lib:pbluescriptII
SK plus clone:hk03632.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS Nagase,T., Ishikawa,K., Suyama,M., Kikuno,R., Miyajima,N.,
Tanaka,A., Kotani,H., Nomura,N. and Ohara,O.
TITLE Prediction of the coding sequences of unidentified human genes. XI.
The complete sequences of 100 new cDNA clones from brain which code
for large proteins in vitro

JOURNAL DNA Res. 5 (5), 277-286 (1998)
MEDLINE 99087487
REFERENCE 2 (bases 1 to 4318)
AUTHORS Ohara,O., Suyama,M., Nagase,T., Ishikawa,K. and Kikuno,R.
TITLE Direct Submission

JOURNAL
Submitted (08-OCT-1998) Osamu Ohara, Kazusa DNA Research Institute
Laboratory of DNA Technology; Yana 153-2, Kisarazu, Chiba
292-0812, Japan (E-mail:cdnainfo@kazusa.or.jp, Tel:+81-438-52-3913
Fax:+81-438-52-3914)

FEATURES	Location/Qualifiers
source	1. .4318

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="hK03632"
/sex="male"
/tissue_type="brain"
/clone_1b="pBluescriptII SK plus"
/dev_stage="adult"
1..3015
/gene="KIAA0730"
<1..3015
D5S

```

BASE COUNT 1395 a 734 c 847 g 1342 t
ORIGIN

Query Match	100.0%;	Score 593;	DB 9;	Length 4318;
Best Local Similarity	100.0%;	Pred. No. 4.7e-92;		
Matches 593;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0

Qy	1	acccttaatgcttaacagagcttcctggtttgtagaagatagaacagggaaactcgaatgct	60
Db	3650	ACACCTTAATGTTTACAGGCTTCCTGTTGTAGTAAGATAGGACAGGAACATCTAAAAATGGT	3700
Qy	61	ggcagctctcttaataccagcttggttagtactgttctcgtgaaactcgtctccacaagaacat	120
Db	3710	GGCAGTCTCTATTTACAGGTGTTAGTATGTTTCTGGAACCTGGTTGCCAAGACACATT	3760
Qy	121	tattaactgttaagaacacctgaccttaagttggttggtgtaacatactttccacaatggtata	180
Db	3770	TATTTAACGTGTAGAACACTGGTTTAAGTTGTGTGTACATATTTTTCCAAAAATGTTATA	3820
Qy	181	attataatagctgtgcttggaaacagatgcgaactcttctgtctcaaaagtgcacgttaa	240
Db	3830	ATTATATAGAGTGCTTGGAACAGGATGCATCTTTTGTGCTRAAGAGCTCTGCAGTTAA	3880
Qy	241	aaaaaaaacaaccttctcttccaatalgycatgtagtggagctttttaaacttaaaac	300
Db	3890	AAAAAAAAACAACCTTTCTTTTCATATAGCATATGCGATGATGAGGAGTTTAAAACTTTAAAAAC	3940
Qy	301	atcaaaaatctttaaactcattggttatcatagtaagctttaattatcaggcttatattc	360
Db	3950	ATCAAAAATTTGTTAAAAATCATGTGTTATCTAGTAGAGTTTATAATTATCGGCTTATATTTC	4000
Qy	361	cccatgaatgatacgaactgacatttaattcatagttgtgtcgcgaatgctctcttaactt	4200

Db	4010	CCCATGAAATGATCAGAACTGCACATTTTATTCANGTTTGTCGCCATGCGTTTACTTCTT	4068
Oy	421	aacatattctctttgcagaaatgaaaaagtaatgataatagtttaataagtgactg	480
Db	4070	AACATATTTCTTTGCAGAAATGTAAAAAGTAATGATTAATTTAGTTATATAACTGTACGG	4128
Oy	481	ctgtaaatgactgctaaatatactttagcaataaaggtctacagaaatgltgaact	540
Db	4130	CTGTAAAGATGCTAAATATATCTTATGCAATTAAGGCTTACAGAACATGTTGAAACTT	4188
Oy	541	ttttactcttattatgggaataagaatglttgacctccaattattatgct	593
Db	4190	TTTTTACTTTTATTTGGAAATAGGAATGTTTGACCTCCACATTTTATTTATGCTT	4242

RESULT 2
AX119931

LOCUS	AX119931	12793 bp	DNA	linear	PAT 11-MAY-2003
DEFINITION	Sequence 1 from Patent WO01292265.				

ACCESSION	AX119931
VERSION	AX119931.1
	GI:14036678

KEYWORDS . human
SOURCE

	Homo sapiens
ORGANISM	
Eukaryota; Metazoa;	Chordata; Vertebrata; Euteleostomi;
Mollusca;	Cnidaria; Mollusca; "Mollusca";
	"Mollusca";

REFERENCE 1 (bases 1 to 12793)
Mammalia; Euteria; Primates; Catarrhini; Hominoidea; Homo
Homo sapiens; Hominidae; Homininae; Hominini; Hominina; Hominid;
Primate; Mammal; Vertebrate; Chordate; Tetrapoda; Amniota; Mammalia;

AUTHORS Hudson, T.O., Engler, C.O. and Richter, A.
TITLE Identification of arsacs mutations and methods of use
JOURNAL Patent: WO 0120265-A1 26-APP-2001.

FEATURES
JOURNAL
FALGUT. MO 0122200 A 1 20 APR 2001,
MCGILL UNIVERSITY (CA) ; Hopital Sainte-Justine (CA)
location/Qualifiers

```
source
1. .12793
/organism="Homo sapiens"
```

BASE COUNT	4163 a	2256 c	2487 g	3887 t
	/db_xref="taxon:9606"			

ORIGIN

Query Match 100.0%; Score 593; DB 6; Length 12793;

Best Local Similarity 100.0%; Pred. No. 3.6e-92;
Matches 593; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 acatcctatgtttacaggtctccgttttgatgaagatagcaacgnaaacctaaatggt 60
|||||
Db 12201 ACATCCTATGTTTACAGGCTTCTCTGTTTGAATGAAGATAGCAACGGAACCTCAAATGTT 12260

QY 61 ggcagttctattaccagttgttagtattgttcttgaacgccttgccaagacaact 120
|||||
Db 12261 GGCAGTTCCTATTACCAAGTGTGTAGTATGTTCTCGGAACTGCTTCGCCAAGCAACAT 12320
|||||

QY .121 tattaactgttagaacactgccttattgttgtgtgtacacatttccacaatgtctara 180
|||||
Db 12321 TATTAAGTGTAGAACACTTGCTTATGTTGTGTGTACATATTTCCACAAATGTTTA 12380
|||||

QY 181 attcatataagtcgtgctgaacagaatgcactctttgttgcctaagaagtgcctcagttca 240
|||||
Dh 12381 ATTCTTATACGTCGCTGC AAGCAAGCATATCATTCTTTCTTGCTTAAGAAGCTCGCTCAGTTAA 12440

[illegible]

QY 301 atcaaaatcgttaaatacatcgtgtacatagtagttataaataatcgcgtatattc 360

361 cccatgaatgatacgaactgacatttaattcatgtttgtctgcgcacgtctcttacttt 420

421 aacatactcttgcgaatgtaaaagtaacgataatagttatataagtgtactgg 480

Db 12621 AACATATTTCTTTCGAGATGTAAGATGATATAGTTATATATAAGTACTGCG 12680
Qy 481 ctgtaaatgctgtaataataacttataatgaaggtctacgaacatgttgaaactt 540
Db 12681 CTGTAAATGATGCTAAATATACCTTATGCAATTAAGGGCTTACGAACATGTGAAACTT 12740
Qy 541 tttctacttactatggaatacaagatgcttgcacccacacatttactgctt 593
Db 12741 TTTTACTTTTATTTGGGAAATAGGAATGTTTGCACCTCCACATTTTATTCCTT 12793

RESULT 3
AF193556 12793 bp DNA linear PRI 07-FEB-2000
LOCUS Homo sapiens saccin (SACS) gene, complete cds.
DEFINITION AF193556
ACCESSION AF193556.1 GI:6907041
VERSION AF193556.1
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 12793)
Engert,J.C., Berube,P., Mercier,J., Dore,C., Lepage,P., Ge,B.,
Bouchard,J.P., Mathieu,J., Melancon,S.B., Schalling,M.,
Lander,E.S., Morgan,K., Hudson,T.J. and Richter,A.
ARSACS, a spastic ataxia common in northeastern Quebec, is caused
by mutations in a new gene encoding an 11.5-Kb ORF
Nat. Genet. 24 (2), 120-125 (2000)
2 (bases 1 to 12793)
Engert,J.C., Berube,P., Dore,C., Lepage,P., Ge,B., Hudson,T.J. and
Richter,A.
Direct Submission
Submitted (08-OCT-1999) Genome Centre, Montreal General Hospital,
1650 Cedar Ave., Montreal, QC H3G 1A4, Canada

TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS
FEATURES
source
1. 12793
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="13"
/map="between D13S232 and D13S292"
1. 12793
/gene="SACS"
/product="saccin"
1. 12793
/gene="SACS"
77. 11566
/gene="SACS"
/note="molecular chaperone"
/codon_start=1
/product="saccin"
/protein_id="AAF31262.1"
/db_xref="GI:6907042"
/translation="MNTFWPGRELIVQWYFEDNRHPSYMLKMYKXNYHIFSEDL
TFDEMRPLRTLEGGQVEELIRIRISVYLDDSEKQDEPLADLYOKAGGYVL
KKLDAIQPLIKTIHSPPLSAVIDIMEMLQKLCNQTSLPRLHAKFVLASL
TSSSEKERKIQLALFKRIHNSDGISSYTKLCKQVLAHNAKLPADRLISVID
SSDEATIRLANMKIEQLTKTSLCLAVLDIENAFVSHETVTOIMLVENLSLNE
NNPVEMLPLPKFIOISQOMYASAGELPFDIVLDECEGTEPFPVPSFDLL
HSRLQGLKNEASLKEKDYVOYAKTEALOVGACPRODYLKAKKLLIYANKNHL
OSSEGMKTIKTIKYPACKERPNYRPGSLWKDLCNLCAPPKMCVGAHILIGSSLP
LVSTHVNLEKALGIFTKPSLAVLNHFKITVDVYSKTFSDYDQFQHILLETIGF
MDHLNEGDSFRALKFPMWWTGKFCPLAQAIVKPIHDLADLPYLANHPTAKHFO
LFRVCSIELTSDHISIMYIOLIKYLSDDQLDSEKQNLHMLNIRMLYSQIPAS
PNTPVPIHHSKNPSKLMKPHECCYCDKVDLDLSDVPIILHVDIMPKTAE
WLKVPCLSTRLLINPMWFGESQREPLRYRIKRLNILEPVSVDIKETLQMDANA
TCSSTLIDMRNMDIRENLDPGMAKCHPRLMSPNNSQPSDEDFPNITRIGESLARK
EVDKVGKFLGNSVYHITDIPITMSREFNIMDPINHSKIKDKMSNGIKIIMNSK
OOKRLKRPNQKPFIDVFCQLPLVEAPVSYNGTILFRISFTQDEAKSEVSTCY
MTADISLVEDSLGCHRLIFITQSVSMYLYKLTETNPSLAOTVILIKKSSCSK
ALNTPLVSLKEAKLMTKTCSSSNKRLPSEPKSCILQITVEEFHVFRIADLSP
LFRGPDDDPAALFEMAKSGQSKKPSDELQKTYECTTWTLLCTGMDTGALKFLSLSSG

RLGLVPCGAVGVLSEIOPDKMTVKPHIGVEVCEPLRLIKTGLPYHINCFVATNSR
KEIMKTDGTGRNMFTFMRHYIVAYLOVLSVLBDLATSGLMDTYTAAVAPDDDLVD
DESVICQGFEDTIAHGKGLTLVFPSDGSWSMKVRPLDDSLTRBDVGSAAKATF
LKYLKTGKSNILCAVELPSSSVKLGFEEDGCKQLILENTSEKQFSEVFPNLOETEA
ELRDPMLPIFVNLKGVDEFSGVLVTPDIPCSLGHVLYLPSRLIHGGRVAKLPJLD
GRPPSYTODYLNPIILIKLVQGMKADLTLMDDMLERAVASVAEINKSDVAALCSS
TILSLDEKTKIRDPRAKDEAAKQYOTIRPLPLPKRPGFSLDMKNSFKRETEMAAT
LYTAEHODIYCILOPLINENSHSPFGCGSVLAIVKEFLGKPKYDVLVNOLEKAV
SVDDGTLVQENITNACTYLHEALMONEITKASIDKLPPSILVENAVYSEKVS
FLNLEAPALYQLPNKYKNNFRELFEVGVQSCVVEDALVESIDQGRKQVITE
ENFOLCRRIISGFIWSLIREKQCEKNGYKILIDPTNMLPLASICYNDPCMIK
KOTTYKACHADIPREVAVKGAIVKRRKALERASVNCFTTLTEFGOKRSLRIS
ILNAYPSEKEMKELLOMADAKATEICFEFDRORHVDRIPDDKAPLOGPRLAVYN
NOPFTEDVRCIGIONLKGTYEGNPKYTGQYIGIFSNVYHITDDEPSIFSDIICIDP
HANYAPATISISPGKMRDIDADRFQFSLVDLYGTHKRLDCTHFRPLRANEA
KVEISSVPASDMVQNLDDKLRSDGELMLPILHMEKISICEDKSTGLVNLVSVK
GKITDGBRLKRPQPHASVIDITLPRGVAACITHNKPKPRAKCFPLSLTEGSLPH
GFSMEKVSYSIAHKNODITLPRGVAACITHNKPKPRAKCFPLSLTEGSLPH
VNGHPALDSARRMLMRDNGVGRSDNNLSMTALAPAVVELLIDQLKRYPPGSDPT
LSYQNTPRHYVVDTLKPLSPFPVNRLDIQPRLCYLKLAVYNCIHEDMKRLPYVA
PNIDGDSHSAVITITWIMSTSKTRFPFONLQDELQHLKMDYNTTKRYTAENVY
RLKHLLEIGFNLVYNCDETANLYHCLIDADIVSVTPADISFLMTFESSPNTCH
GKLPCRLOQTNLKLHFSKLVDVCFEDAEENEIEVEGLPLITLDSVLOTPEAKRK
FLTYHELIPSRDLEFNTLYLKYSNILNCKYAKVADISSFADLSSVLPREKTKS
CTKKONFASSESLKNAWHFISSESVYKEDQEEFKPTFOIVDTLQMDALPGKFTV
SANQLYVPBQDVLPLSLMIANFPAKNSKVFHALMKACIOLALNKTSKSAVP
LISCHTANIESPTSLKALHMYOTSTFRAEKLEVENDEALMIFNCLNHLMSODI
KILKSLPCKYSISGRVYSIGKFGCYVLTISPSAEKQFROLSAELSLIKHLEL
YEVIGAPVDLDEVLYLHLLPKILENSYDALKLHLLPKIRLSAELSLIKHLEL
LESLLITHDNSRLKOKHFDYRTVRYVEVMEPEKLEIPDRFKLQDLKPKNHVF
MTSVBEFLRNIGLKYILISQOOLQFAKEISVANTENWKSLETONTQDILNHFQOR
MDLSGFKLELSLIPLCPEBRAPEIRRHPOYQVNGTLPILKFGNAYLNHFKROC
DVLQMLTSCPIPERKATPLSIEQESDLPQBOQLVNLMLVNLDPDLVYINNC
RNICNTITLDEEMVKTRAKVLRSLYEELSEKKEFRQLGVAFVAVMEDKMLKPE
VVNLNYESDPKRYLYKLPLELCTFHLFSEKLEDEIDISFKYVEVLSRIFKSEGO
LDENKMTYRVAVSGLERSLONDSVAKRSLDPLATLPQDGGFLVSSILFQDE
APHYKRIQNGINQVOMLYDSOCYGLDHQFPHKRLMLPQKLRPLLSLLEQDE
ETPRKVCQFGLCSLQGRQLDLSSEDTITLIRIMKENDNATLANEKAIRLCKALR
EGLKVSCEFERLQTTLRVKGFPNIPHSSEFPAKLRGNAVILLYIQHSQSKLEIFL
ALAMTKLSANDNLISDYSILAMIGCDNIRIEBKLDKSLVSKSLLEIPMT
PIPAEILYTLMDPMNVYFGEVYGVLDVDEGDIIGSYOPTYATILIOVEVEDAD
NSSFLIKTYOIGDISEYKIVSSLDLYKFRPSESSQSRSPATPSTYPIPEPLTGR
SIRPLPSGRSHKTSKSHQSPKTLKVNLSLEILKEYSVYQVAKMLPESRKKIIRTL
YLKHPDKNPENDIANEVKRLHONELKONNEKCPKPAQGYVSOQFVPPPKSVGNVER
KYSFQRYTYSNMOEATSHKSEKQONNEKCPKPAQGYVSOQFVPPPKSVGNVER
KMLROARANFSAAARNDLHKNNANMVCFCYSLKRLAIADYAVRGSDVDVPTALA
OKLTSEYSOOLEGLNDVHTLEAVGVDSLKRYPDLPLFPQIPMDRFTSVSAVMRECT
ACIIIKLENMOOKV"

BASE COUNT * 4163 a 2256 c 2487 g 3887 t
ORIGIN

Query Match 100.0%; Score 593; DB 9; Length 12793;
Best Local Similarity 100.0%; Pred. No. 3 6e-92;
Matches 593; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 acatctatgtttacaggtctctcgtttgatgaagatagcaacggaactcaaatgt 60
Db 12201 ACATCTTATGTTTACAGGCTCTCTGTTTGAAGATAGCAACGGAAACCTCAAAATGCT 12260
Qy 61 ggcagttcttatccagttgttagtatgtttcttggaactgtttgcgaagaactt 120
Db 12261 GGCAGTCTTATTCACGAGTTGTAGTATGTTTTCGAAACCTCTGCAACACAACTT 12320
Qy 121 tattaactgttagaaccttgccttatgtttgtgtacacatttccacaactgttata 180
Db 12321 TATTAACTGTAGAAACTTCTTATGTTGTGTGTAACAATTTTCCACAAATGTTATA 12380
Qy 181 attataagtggttggaacaggaatcaactcttggttgtctaaagtgctgaagtaa 240
Db 12381 ATTATATAGTGTGTTGAACAGATCAATCTTTTGTGCTTAAAGGTCTCGAGTTAA 12440
Qy 241 aaaaaaaaaaaccttctctcctaataatgagcatgagtgaggtcttctttaaacttaaaaaac 300

```

|||||
Db 12441 AAAAAAAAAACCTTTCTTCAATATGCGATGAGCGTTTCTTTAACTTTAAAC 12500
QY 301 atcaaaatgttaaaatcatgttattatagttatattatcggcttatattc 360
Db 12501 ATCAAAATATTGTTAAATCAATGTTATCTAGTTATTAATATCGCGCTTATATTTC 12560
QY 361 cccatgaatgatcagaactacattcaatcagttgtctgcgcacgtctcttactt 420
Db 12561 CCCATGATGATCAGAACTGACATTAATCATGTTTGCTCGCAGCTCTTTACTTT 12620
QY 421 aacatatctcttgcagaatgtaaaagtgatgataattatgatatagtgctag 480
Db 12621 AACATATTCTTTTGCAGAAATGTAAGGTAATGATTAATTAATTAAGTACTCG 12680
QY 481 ctgtaaatgctgaataactattatgcaattgaagggcttacagaaacgttgaact 540
Db 12681 CTGTAATGATGCTAAATATCTTTATGCAATTGAAGGGCTTACAGAACTGTGAACCT 12740
QY 541 ttcttacttattgggaataagaatgttgcacctccacatttacttct 593
Db 12741 TTTTACTTTTATTTGGAATAAGAAATGTTTGCACTCCACATTTATTTCTT 12793

RESULT 4
AL157766 92693 bp DNA linear PRI 11-APR-2001
LOCUS AL157766/c
DEFINITION Human DNA sequence from clone RP11-40020 on chromosome
ACCESSION AL157766
VERSION AL157766
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 92693)
REFERENCE
AUTHORS Tromans,A.
TITLE Direct Submission
JOURNAL Submitted (11-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
COMMENT
requests: clonerequest@sanger.ac.uk
On Apr 12, 2001 this sequence version replaced gi:12709868.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em., EMBL; Sw.,
SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 13, constructed by the Sanger Centre Chromosome 13
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr13
RP11-40020 is from the library RPc1-11.1 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6
IMPORTANT: This sequence is not the entire insert of clone
RP11-40020. It may be shorter because we sequence overlapping
sections only once, except for a 100 base overlap.
The true left end of clone RP11-760M1 is at 92594 in this sequence.

```

```

FEATURES             The true right end of clone RP11-72P19 is at 100 in this sequence.
source               Location/Qualifiers
                    1..92693
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /chromosome="13"
                    /map="q12.11-12.2"
                    /clone="RP11-40020"
                    /clone_1lb="RPc1-11.1"
                    2390..2485
                    /note="MIR repeat: matches 81..192 of consensus"
                    2562..2673
                    /note="MIR repeat: matches 68..183 of consensus"
                    3896..4201
                    /note="AluY repeat: matches 3..308 of consensus"
                    5122..5397
                    /note="MER46C repeat: matches 1..286 of consensus"
                    18986..19294
                    /note="AluY repeat: matches 1..310 of consensus"
                    19644..19873
                    /note="MER46A repeat: matches 1..235 of consensus"
                    20613..20912
                    /note="AluSg1 repeat: matches 1..300 of consensus"
                    23342..23651
                    /note="AluSg1 repeat: matches 1..309 of consensus"
                    24769..24891
                    /note="L2 repeat: matches 2554..2682 of consensus"
                    25871..26011
                    /note="L2 repeat: matches 2356..2495 of consensus"
                    26033..26109
                    /note="L2 repeat: matches 2601..2688 of consensus"
                    26245..26344
                    /note="L2 repeat: matches 2154..2255 of consensus"
                    26938..27096
                    /note="MIR repeat: matches 3..175 of consensus"
                    27150..27653
                    /note="L2 repeat: matches 1063..1644 of consensus"
                    28522..28891
                    /note="THE1B repeat: matches 1..364 of consensus"
                    29447..29834
                    /note="LIMB repeat: matches 5787..6164 of consensus"
                    36098..36415
                    /note="AluSx repeat: matches 1..308 of consensus"
                    37202..37414
                    /note="MIR repeat: matches 22..262 of consensus"
                    37963..38254
                    /note="AluSg repeat: matches 9..301 of consensus"
                    38703..39008
                    /note="AluSg repeat: matches 1..306 of consensus"
                    39790..40093
                    /note="AluSx repeat: matches 1..304 of consensus"
                    40126..40416
                    /note="AluSg repeat: matches 1..292 of consensus"
                    40444..40733
                    /note="AluSg repeat: matches 1..292 of consensus"
                    41322..41405
                    /note="Single clone region. Assembly confirmed by
                    restriction digest data."
                    41541..41788
                    /note="AluSg repeat: matches 1..248 of consensus"
                    44790..45101
                    /note="AluSg repeat: matches 1..313 of consensus"
                    45261..45312
                    /note="13 copies 4 mer tggc 88% conserved"
                    45899..46206
                    /note="AluY repeat: matches 1..307 of consensus"
                    46754..47052
                    /note="AluY repeat: matches 1..298 of consensus"
                    47067..47365
                    /note="AluY repeat: matches 1..299 of consensus"
                    47477..47873
                    /note="L1MA10 repeat: matches 5950..6322 of consensus"
                    47889..48229

```

```

repeat_region /note="AlusX repeat: matches 1. .312 of consensus"
49168. 49212 /note="Alu repeat: matches 85. .126 of consensus"
repeat_region /note="L2 repeat: matches 1685. .1757 of consensus"
50704. 51032 /note="AlusX repeat: matches 1. .308 of consensus"
misc_feature 52204. 53009 /note="CPG Island"
repeat_region /evidence=not_experimental
53978. 54137 /note="AlusX repeat: matches 1. .160 of consensus"
54179. 54511 /note="L1PB2 repeat: matches 5822. .6172 of consensus"
55685. 55949 /note="AlusX repeat: matches 1. .305 of consensus"
57331. 57390 /note="L2 repeat: matches 1685. .1757 of consensus"
57357. 57392 /note="AlusX repeat: matches 1. .305 of consensus"
57753. 57930 /note="L2 repeat: matches 2581. .2656 of consensus"
58260. 58389 /note="MIR repeat: matches 82. .262 of consensus"
58564. 58611 /note="MIR repeat: matches 2. .153 of consensus"
59350. 59533 /note="L2 repeat: matches 261. .294 of consensus"
59350. 59533 /note="AlusX repeat: matches 129. .313 of consensus"
59992. 60223 /note="Alu repeat: matches 85. .299 of consensus"
61036. 61144 /note="L2 repeat: matches 2581. .2656 of consensus"
62008. 62187 /note="L2 repeat: matches 2238. .2418 of consensus"
62188. 62316 /note="Alu repeat: matches 1. .129 of consensus"
62330. 62363 /note="Alu repeat: matches 261. .294 of consensus"
62362. 62565 /note="L1PB2 repeat: matches 1586. .1787 of consensus"
62566. 62865 /note="AlusX repeat: matches 1. .302 of consensus"
62866. 64385 /note="AlusX repeat: matches 1. .302 of consensus"
64386. 64694 /note="L1PB2 repeat: matches 46. .1586 of consensus"
64695. 64713 /note="Alu repeat: matches 1. .306 of consensus"
65068. 65395 /note="L1PB2 repeat: matches 29. .46 of consensus"
65396. 65569 /note="L1PB2 repeat: matches 5405. .5733 of consensus"
65571. 65640 /note="Alu repeat: matches 136. .309 of consensus"
65696. 65717 /note="L1PB2 repeat: matches 5728. .5791 of consensus"
65723. 66096 /note="L1PB2 repeat: matches 5789. .6155 of consensus"
66371. 66410 /note="L1PB2 repeat: matches 5789. .6155 of consensus"
67586. 67886 /note="L1PB2 repeat: matches 5789. .6155 of consensus"
69748. 69930 /note="Alu repeat: matches 1. .299 of consensus"
70957. 71267 /note="MIR repeat: matches 6. .248 of consensus"
71279. 71413 /note="Alu repeat: matches 1. .311 of consensus"
71411. 71737 /note="MIR repeat: matches 548. .680 of consensus"
71780. 72075 /note="MIR repeat: matches 47. .485 of consensus"
72145. 72256 /note="AlusX repeat: matches 1. .295 of consensus"

```

```

repeat_region /note="MER31-internal repeat: matches 42. .175 of consensus"
72454. 72865 /note="MER31-internal repeat: matches 332. .739 of consensus"
repeat_region /note="MER31-internal repeat: matches 883. .1261 of consensus"
72873. 73249 /note="MER31-internal repeat: matches 883. .1261 of consensus"

Query Match 100.0%; Score 593; DB 9; Length 92693;
Best Local Similarity 100.0%; Pred. No. 2.2e-92;
Matches 593; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 aacatctatggttcacagcttcctgcttgcagagatagacagagaaactcaaatgct 60
Db 6709 ACATCTTATGTTTACAGGCTTCTGTTGATGAGATGACAGGAAACTCAAAATGCT 6650
QY 61 ggcagcttctatcacagcttgcagcttgcagcttgcagcttgcagcttgcagcttgc 120
Db 6649 GGCAGTCTTATTCAGAGTCTTGTATGTTGTTTGCAGAACTGCTGCAGACACAT 6530
QY 121 tattaactgttagaacactgtcttgcagcttgcagcttgcagcttgcagcttgcagct 180
Db 6589 TATTACTGTAGAACACTTCTTATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 6530
QY 181 attatagtggttgcagagatgcagcttgcagcttgcagcttgcagcttgcagcttgc 240
Db 6529 ATTATATAGTGTGTTGAGACAGATGCAATCTTTGTTGTTGTTGTTGTTGTTGTT 6470
QY 241 aaaaaaacacaccccttcttccaaatgcagcttgcagcttgcagcttgcagcttgcag 300
Db 6469 AAAAAAACACACCTTCTTCCAAATGCGATGAGTGAAGTGAAGTGAAGTGAAGTGA 6410
QY 301 atcaaaaatgttaaaatcaatcgttgcagcttgcagcttgcagcttgcagcttgcagct 360
Db 6409 ATCAAAAATGTAAATCAATCAGTGTATGTTGTTGTTGTTGTTGTTGTTGTTGTT 6350
QY 361 cccatgatgctgcagactgcttgcagcttgcagcttgcagcttgcagcttgcagcttgc 420
Db 6349 CCCATGATGCTGCAGACTGCTTGCAGCTTGCAGCTTGCAGCTTGCAGCTTGCAGCTT 6290
QY 421 aacatcttcttgcagactgcttgcagactgcttgcagactgcttgcagactgcttgcag 480
Db 6289 AACATCTTCTTGCAGACTGCTTGCAGACTGCTTGCAGACTGCTTGCAGACTGCTTGC 6230
QY 481 ctgtaaatgctcaaatatcattatgcagcttgcagcttgcagcttgcagcttgcagct 540
Db 6229 CTGTAATGCTCAAAATATCTTATGCAATTAAGGCTTGCAGACTGCTTGCAGACT 6170
QY 541 ttctacttattggaataagaaatgcttgcagcttgcagcttgcagcttgcagcttgc 593
Db 6169 TTTTACTTTTATGGAATAGGAATGCTTGCAGCTTGCAGCTTGCAGCTTGCAGCTT 6117

RESULT 5
AC079761 99819 bp DNA linear HTG 10-SEP-2000
LOCUS AC079761
DEFINITION Homo sapiens chromosome UNK clone R11-143617, *** SEQUENCING IN
ACCESSION AC079761
VERSION AC079761.1 GI:10047966
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 99819)
AUTHORS Waterson, R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 99819)
AUTHORS Waterson, R.H.
TITLE Direct Submission
JOURNAL Submitted (10-SEP-2000) Genome Sequencing Center, Washington

```



```

misc_feature      /note="assembly_name:Contig57"
28879..30893
/note="assembly_name:Contig58"
misc_feature      30994..32460
/note="assembly_name:Contig59"
misc_feature      33561..33984
/note="assembly_name:Contig60"
misc_feature      34085..35285
/note="assembly_name:Contig61"
misc_feature      35386..37184
/note="assembly_name:Contig62"
misc_feature      37285..39172
/note="assembly_name:Contig63"
misc_feature      39273..40874
/note="assembly_name:Contig64"
misc_feature      40975..42893
/note="assembly_name:Contig65"
misc_feature      42994..44384
/note="assembly_name:Contig66"
misc_feature      44485..45999
/note="assembly_name:Contig67"
misc_feature      46100..48669
/note="assembly_name:Contig68"
misc_feature      48770..50798
/note="assembly_name:Contig69"
misc_feature      50899..52809
/note="assembly_name:Contig70"
misc_feature      52910..55127
/note="assembly_name:Contig71"
misc_feature      55228..58087
/note="assembly_name:Contig72"
misc_feature      58188..61004
/note="assembly_name:Contig73"
misc_feature      61105..64185
/note="assembly_name:Contig74"
misc_feature      64286..67105

Query Match      100.0%; Score 593; DB 2; Length 99819;
Best Local Similarity 100.0%; Pred.No. 2,2e-92;
Matches 593; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      481      ctgtaaatgtagtcaataatactatcatgcaattaaagggtcacaagaacatgtgaaact 540
|||||
Db      61602  CTGTAATGATGCTAAATATACCTTTATGCAATTAAAGGCTTACGAACATGTTAAACTT 61661
|||||

QY      541      ttcttactttatgtggaaataagaatgttgcaccctccacatttattgctt 593
|||||
Db      61662  TTTTACTTTTATTTGGGAATAGCAATGTTGCACCTCCACATTTTATTCGCTT 61714
|||||

```

```

RESULT      6
AB056815      3289 bp      mRNA      linear      PRI 14-MAR-2001
LOCUS      Macaca fascicularis brain cDNA clone:Qf1A-15307, full insert
DEFINITION
sequence.
AB056815      GI:13365931
AB056815.1
VERSION      f1s (full insert sequence): oligo capping.
KEYWORDS      Macaca fascicularis adult male frontal lobe left cDNA to mRNA,
SOURCE      clone.lib:macaque brain cDNA library Qf1A clone:Qf1A-15307.
ORGANISM      Macaca fascicularis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
Cercopithecinae; Macaca.

```

```

REFERENCE      1 (sites)
AUTHORS      Osada,N., Hida,M., Kusuda,J., Tanuma,R., Iseki,K., Hirai,M.,
Terao,K., Suzuki,Y., Sugano,S. and Hashimoto,K.
TITLE      Isolation of full-length cDNA clones from macaque brain cDNA
JOURNAL      Unpublished
AUTHORS      2 (bases 1 to 3289)
TITLE      Hashimoto,K., Osada,N., Hida,M., Kusuda,J. and Sugano,S.
JOURNAL      Direct Submission
AUTHORS      Submitted (09-MAR-2001) Katsuyuki Hashimoto, National Institute of
TITLE      Infectious Diseases, Division of Genetic Resources; 23-1, Toyama
JOURNAL      1-chome, Shinjuku-ku, Tokyo 162-8640, Japan
AUTHORS      (E-mail:khashi@nih.go.jp, URL:http://www.nih.go.jp/yoken/genbank/,
TITLE      tel:81-3-5285-1111(ex.2120), fax:81-3-5285-1181)
JOURNAL      Lab host: TOP10

```

```

COMMENT      Vector: pME18S-FL3 (Acc.No. AB009864)
R. Site1: DraIII (CACTGTG)
R. Site2: DraIII (CACCATG)
Description: 1st strand cDNA was primed with an oligo(dT) primer
using specific 5' and 3' primers and amplified by PCR. The PCR
product was digested with SfiI and size selection was performed to
exclude fragments <1.5kb. The SfiI-digested PCR product was cloned
into distinct DraIII sites of pME18S-FL3. XhoI sites just outside
the DraIII sites can be used to isolate the cDNA insert. Libraries
were constructed by Sugano et al. (University of Tokyo, Institute of
Medical Science). Custom primer used for sequencing
('5' end primer (CTTGTGCTTAAGCTGCG);
3' end primer (CGACCTGACGTCGACACA) ).
FEATURES
source
1..3289
Location/Qualifiers
/organism="Macaca fascicularis"
/db_xref="taxon:9541"
/clone="Qf1A-15307"
/sex="male"
/tissue_type="frontal lobe left"
/clone.lib="macaque brain cDNA library Qf1A"
/dev_stage="adult"
/note="Host: TOP10, Vector: pME18S-FL3 (Acc.No. AB009864);
R. Site1: DraIII (CACTGTG); R. Site2: DraIII
(CACCATG)"
35..1951
/codon_start=1
/product="hypothetical protein"
/protein_id="BAB39340.1"
/db_xref="GI:13365932"
/translaton="MLVDLSQCYLGDHGFHTKLMLFPQKLRPLRLSSILIEBOLDE
TPKVCQFGLCSLQGRQLDSSEQFTTGIRLIMKHENDNAFLANERKARLCIKARE
GLKVSCEKLTOTLRVKGFPDIPHSSEETGAPAFKRGNAVILLYIOHSDSKINFLFLA
LMTLKSATDNLISDTSYLLAMLGCDINDYIIGKLDLSLGVKYDSSESKLELMPGP

```

CDS

```

QY      361      ccacatgaatgacagaactacaattcaatcatgtgtctcgacagctcttactctt 420
|||||
Db      61482  CCATGATGATACGAACGTACATTATTCATGCTTGTCTCCGCAATGCTTCTTACTTT 61541
|||||

QY      421      aacatattctttgcagaaatgaaagtaatgataatagattatataagtgactgg 480
|||||
Db      61542  AACATATTCTTTGCAGAAATGAAAGTAATGATTAATAGTTTATATAGTCTACTGG 61601
|||||

```


BASE COUNT 1062 a 552 c 653 g 1022 t

ORIGIN

1PAETHYLLMDPMNVPEGVYLVDAEGDYGSDYPTTYTAIVOEVEDADN
SSFLGIYQIDIGYSEKIVSSLDYKFSRPSQSDSADSTPSPFLPGLRS
IPFLGRESHSTSSKHOSPKIKLVNLEPEIKVEYVVEOMAKLPESPKEKILRLY
LKMHPDNENNDIANEFKHONENREKOFADLONDRASRRPESTASPOSOK
YSFORPTSMNOBATSXKSEKROOKKEKPPSAGQYSGRPFPYPFKSYGNVEAR
WIRQANFSAARNDLKKNANWVCKYLSKTLALVADYAVRAGSDSDVKTALAO
KLEESQLEGLTNDVHTLEAYGVDSLKTRYDLPFPDIPNDRFSEVAMRYECTA
CIIRKLENFOOKV"

Query Match 71.9%; Score 426.6; DB 9; Length 3289;
Best Local Similarity 89.6%; Pred. No. 1e-63;
Matches 544; Conservative 0; Mismatches 39; Indels 24; Gaps 7;

QY 1 acatcttaagttaacaggtctctgtttgataagataagcaacgaaactcaaatgtc 60
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
DB 2589 ACATCTTGTGTACAGGCTTCCTGTTCAATGAGATGCAATGCAAAACCTAAATGCT 2648
QY 61 ggcagttcttataccagttgttagtattgt---ttctgaaactgctgcaagacaa 116
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
DB 2649 GGCAGTCTTATTACAGATGAGATGTTAGTATTGTTCTGGAACGCTGCGCAAGCAA 2708
QY 117 cattatlaactgtagaacactgcttctatgltgtgtgtacatatcttccacaatgt 176
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
DB 2709 CATTATTATACGTTAGAACACTGCTTTATGTTGTGTATATATTTCACCAATATGT 2768
QY 177 tataattatatagtgtgtgtgaacagatgcaactcttctgtcttaagtgctgcag 236
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
DB 2769 TATTAATTTATATAGTGTGTGAACAGATGCAATCTTTGTCTTAAAGGTGCTGCAG 2828
QY 237 tt-aaaaaaacaaaccccttcttccaatatgcatgtatgtgag-----ttttttta 290
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
DB 2829 TTCAAAAAAATAAACCTTTCTTCAATATGCACTTAGTGACGTTTCTTTTAA 2888
QY 291 ctttaaaacatacaaaattgttaaatcatgtgtatcatagtagtataatlcag 350
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
DB 2889 CTTTAAAAACATTAAACATTGTTAAATCATGTGCTATCTAGAGCTTAAATGATGACG 3448
QY 351 ctatatcttccatgaatgatacagaactgacatttaattcaatgcttctgcacatgct 410
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
DB 2949 CTTATATTTCTTC---ATGTCAGAACTGACATTAAATCACTTTGTCACACATAC 3004
QY 411 tcttcttcaatacatcttcttcttgcagaatgta---aaagtaaatgaatgattta 466
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
DB 3005 T-----CTGTAACATAGTCTTGTGCGCAGATGTTAAAGAAAGTAATGATTAATGTTTA 3059
QY 467 tataagtgtaacgctgtaaaatgataataataacttataatgaatgaaggtctacaga 526
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
DB 3060 TGTAAAGTGTACTGCTGTAATGATGCTAAATATTTATATGCAATTAAGCGCTTACAGA 3119
QY 527 acatgttgaacttttttcttcttcttcttcttcttcttcttcttcttcttcttctt 586
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
DB 3120 ACATGTGTAACAC-TTTTATTAATTATACAGTAATAAGAAAGTTTGCACCTCCACCTGT 3178
QY 587 atctgctt 593
|||||
DB 3179 ATTGCTT 3185

RESULT 7
LOCUS G36555/c 418 bp mRNA linear STS 31-DEC-1997
DEFINITION SHGC-53325 Human Homo sapiens STS cDNA, sequence tagged site.
ACCESSION G36555
VERSION G36555.1 GI:2734222
KEYWORDS STS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 418)

AUTHORS Myers, R.M.
TITLE Human STS (1997)
JOURNAL Unpublished
COMMENT

Contact: Richard M. Myers
Stanford Human Genome Center (SHGC)
Stanford University School of Medicine
Department of Genetics, M-344, Stanford, CA 94305, USA
Tel: 4157259687
Fax: 4157259689
Email: myers@shgc.stanford.edu
Primer A: GTGAGGTCGCAACATTCCT
Primer B: ACATTAAATTCATGTTGTCCCG
STS size: 202
PCR Profile:
Initial incubation: 95 degrees C for 10 minutes
Denaturation: 94 degrees C for 30 seconds
Annealing: 60 degrees C for 30 seconds
Polymerization: 72 degrees C for 23 seconds
PCR Cycles: 30
Thermal Cycler: Perkin Elmer 9600
Protocol:
Template: 25 ng
Primer: each 1 uM
dNTPs: each 200 uM
AmpliTaq Gold Polymerase: 0.07 units/uL
Total Vol: 5 uL

Buffer:
MgCl2: 2.5 mM
KCl: 50 mM
Tris-HCl: 10 mM
pH: 8.3

FEATURES
Location/Qualifiers
1..418
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="13"
/clone_lib="Human"

STS
primer_bind 89..290
primer_bind 89..108
Base COUNT 167 a 65 c 55 g 130 t 1 others

Query Match 55.2%; Score 327.4; DB 11; Length 418;
Best Local Similarity 99.1%; Pred. No. 1.3e-46;
Matches 339; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 253 cttcttcaataatgcatgtatgtgagcttttttaactttaaacaacataaatgtc 312
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
DB 418 CTTTCTTCAATATGCAATGATGAGTGTGTTTAACTTAAACATCAAAATTTGT 359
QY 313 taataatcatgtgtatctatgtagtataataatgagcttatattcccatgaatgat 372
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
DB 358 TAAATCATGTTGTATCTAGTAGTATTAATTAATGCGCTTAATATTCCTCCCATGAATGAT 299
QY 373 cagaactgacatctaactatgcttctgcagatgcttcttcttcttcttcttcttctt 432
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
DB 298 CAGAGCTGACATTTAATCAATGTTTGTCCGCCATGCTTCTTAATCAATATTTCTT 239
QY 433 ttgcagaatgtaaaaggtaatgaatgaatgaatgaatgaatgaatgaatgaatgaatg 492
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
DB 238 TTGCAGAGTGTAAAGGTAATGATTAATTAATTAATTAATTAATTAATTAATTAATGATG 179
QY 493 ctaataatctattatgc-aattaaaggcttaacgaataatgtgaacttttttactttt 551
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
DB 178 CTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 119

OY	310	tgttaaatccttctgtagatagttattataatcacgcctatatattccccaat	369
Db	28528	ATAATATAATATTATGTAATTTTCTGGTTATTTTAAATTTTAA	28587
OY	370	gatcaaacacatttaattcatcttgctcgccatgccttcacttaacaattc	429
Db	28588	TACCATATTTATCTTTTTTCATGATGGTGTTCTATATTTTTTTTTTTCGATTTTTT	28647
OY	430	ccttc 434	
Db	28648	TTTTT 28652	
RESULT	9		
AC004709		196149 bp DNA linear HTG-01-APR-1999	
LOCUS		Plasmodium falciparum chromosome 12, *** SEQUENCING IN PROGRESS	
DEFINITION		***, 3 unordered pieces.	
ACCESSION	AC004709		
VERSION	AC004709.3	GI:4558585	
KEYWORDS	HTG; PHASEL;		
SOURCE	malaria parasite P. falciparum.		
ORGANISM	Plasmodium falciparum		
REFERENCE	Eukaryote, Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
AUTHORS	Hyman,R.W., Fung,E.L., Qin,F., Tamaki,T., Kurd,I.O.B., Conway,A.B. and Davis,R.W.		
TITLE	Plasmodium falciparum 3D7 chromosome 12		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 196149)		
AUTHORS	Hyman,R.W., Qin,F., Fung,E.L., Conway,A.B. and Davis,R.W.		
TITLE	Direct Submission		
JOURNAL	Submitted (21-MAY-1998) Stanford DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA		
COMMENT	On Apr 2, 1999 this sequence version replaced gi:4337173. * NOTE: This is a 'working draft' sequence. It currently * consists of 3 contigs. The true order of the pieces * is not known and their order in this sequence record is * arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence * as soon as it is available and the accession number will * be preserved. * * 1 47631: contig of 47631 bp in length * * 47632 47831: gap of unknown length * * 179125: contig of 131288 bp in length * * 179130 179329: gap of unknown length * * 179330 196149: contig of 16820 bp in length. Location/Organisms 1. 196149 /organism="Plasmodium falciparum" /db_xref="taxon:5833" /chromosome="12"		
BASE COUNT	80057 a 19753 c 18800 g 77138 t	401 others	
ORIGIN	/chromosome="12"		
FEATURES	source		
Query Match	10.3%; Score 61; DB 2; Length 196149;		
Best Local Similarity	47.9%; Pred. No. 0.063;		
Matches 175; Conservative	0; Mismatches 190; Indels 0; Gaps 0;		
OY	70	tattaccagtgttagtatcttgtcttcgtgaacctgtgccaaagaacattatracg	129
Db	119329	TATAACTGTGTTATATATATTTCTAACAAATAGTAGCACTGCTCAATATATATTAATGCTAAAGT	119388
OY	130	ttaagaacctgcctttaactgtcttgctgacatattctccaaagtataatatala	189
Db	119389	TTGGAAATTTTAAATATATTTCTTTTAAATTTGTTTTAAAAAATTTTATATTTCTCTT	119448
OY	190	gtgtgtgtgaacaggatgcactcttgctgctctaaggtcctgcagttaaaaaaaac	249

RESULT	10	
LOCUS	DMU11584/c	
DEFINITION	DMU11584	4601 bp DNA linear INV 23-JUL-1994
ACCESSION	U11584	Drosophila melanogaster Oregon-R mitochondrial A+T region.
VERSION	U11584.1	GI:508826
KEYWORDS	mitochondrial DNA; A+T region; tandem repeats.	
SOURCE	fruit fly.	
ORGANISM	mitochondrion Drosophila melanogaster	

FEATURES	Location/Qualifiers
source	1. .4601

```

/organism="Drosophila melanogaster"
/organelle="mitochondrion"
/strain="Oregon-R"
/db_xref="taxon:7227"
/dev_stage="embryo"
650. .1022
/note="repeat I-A"
/rpt_type= tandem
1023. .1360
/note="repeat I-B1"
/rpt_type= tandem
1361. .1705
/note="repeat I-C/A"
/rpt_type= tandem
1706. .2043
/note="repeat I-B2"
/rpt_type= tandem
2044. .2388
/note="repeat I-C"
/rpt_type= tandem
2491. .2511
/note="deoxythymidylate stretch"
2512. .2648
/partial
/rpt_type= tandem
2649. .3112
/note="repeat II-A"
/rpt_type= tandem

```

Query Match	10.1%;	Score 59.6;	DB 3;	Length 4601;
Best Local Similarity	45.2%;	Pred. No. 0.27;		
Matches 218; Conservative	0;	Mismatches 264;	Indels 0;	Gaps 0;

OY	71	attaccagcttgtagatctgtcttcgggaacgcctcgcaagaacaacttaacgct	130
Db	685	ATTATTATTATATTATTATTATTCGATATATTTAATTAATTAATCAAAATATTATTATTATTA	626
OY	131	tagaacactgccttaagctgtgctgtacaatacttcacasaatgtaactatcatag	190
Db	625	TATATATTGATTTTTAAATAAATTATATTCCTTTTTTATTATTCATATAAAAAATTTTTAAATTA	566
OY	191	tgtggttgacaggaatgcacactctctgtctctaaagctgcgcagcttaaaaaaaca	250
Db	565	TTTTTTTTTAAATAGATTAAATTTTCATTTTTTGTATAAAAAAATATTTTTATATATTAATTT	506
OY	231	acccttcttccaatagtcgacatgtagtggagctttcttaactttaaaaaacatcaaaaatc	310
Db	505	TCAATTTTTTTTTAAAAAAAATTTTTTTAAATTTAATTAATTAATTTTTTTAAAAAAT	445
OY	311	gttaaatcatgtgttactctagtagcttaaatatcgccttactatccccaatgaatg	370
Db	445	TGTAAAAAAAAGTTTAAAAACTCTTAATAAAAAAAATTTTTTTTAGTGATTTAAAAAGAA	386
OY	371	atcagaactgcacttaactcaatgaatgttctgcgcagctcttacttaacatatctc	430
Db	385	TTAAATCTTAATTTTTTAAAAATTTTTTTAAAAAAAATTTTTTTTAAAAAAAATTTTTT	326
OY	431	ttttgcagaatgtaaaagtaatgtaacttaagtatactaaagtgacccggtctgaaaga	490
Db	325	TTTTTAAAAAAAATATTTTTTTAAAAAAGTGAAATTTAAAAATTTTTTTAATTTTAAAAATTA	266
OY	491	tgtcaaatataacttatgcaattaaaggctctacagaaacatgtgaaacttttctacttc	550
Db	265	AAATTTATTTATTTAATATATTTGATTAATATTTTTTTGTAAAAAATATTTTTTTTTTTT	206
OY	551	ta 552	
Db	205	AA 204	

RESULT	11
DMU37541/c	
LOCUS	DMU37541 19517 bp DNA circular INV 02-MAR-2001
DEFINITION	Drosophila melanogaster complete mitochondrial genome.
ACCESSION	U37541
VERSION	U37541.1 GI:1166529
KEYWORDS	.
SOURCE	fruit fly.
ORGANISM	mitochondrion Drosophila melanogaster
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE	1 (bases 12511 to 12682)
AUTHORS	Clary D.O., Goddard,J.M., Martin,S.C., Faucon,C.M. and
	Wolstenholme,D.R.
TITLE	Drosophila mitochondrial DNA: a novel gene order
JOURNAL	Nucleic Acids Res. 10 (21), 6619-6637 (1982)

```

MEDLINE      83090428
REFERENCE    2 (bases 5269 to 5695)
AUTHORS     Clary,D.O., Wahlertner,J.A. and Wolstenholme,D.R.
TITLE       Transfer RNA genes in Drosophila mitochondrial DNA: related 5'
             flanking sequences and comparisons to mammalian mitochondrial tRNA
             genes
JOURNAL     Nucleic Acids Res. 11 (8), 2411-2425 (1983)
MEDLINE     83220794
AUTHORS     3 (bases 404 to 5272)
             de Bruijn,M.H.
TITLE       Drosophila melanogaster mitochondrial DNA, a novel organization and
             genetic code
JOURNAL     Nature 304 (5923), 234-241 (1983)
MEDLINE     83245048
AUTHORS     4 (bases 804 to 1778)
             Satta,Y., Ishiwa,H. and Chigusa,S.I.
TITLE       Analysis of nucleotide substitutions of mitochondrial DNAs in
             Drosophila melanogaster and its sibling species
JOURNAL     Mol. Biol. Evol. 4 (6), 638-650 (1987)
MEDLINE     88174373
AUTHORS     5 (bases 5268 to 13619)
             Garesse,R.
TITLE       Drosophila melanogaster mitochondrial DNA: gene organization and
             evolutionary considerations
JOURNAL     Genetics 118 (4), 649-663 (1988)
MEDLINE     88212147
AUTHORS     6 (bases 441 to 2967)
             Satta,Y. and Takahata,N.
TITLE       Evolution of Drosophila mitochondrial DNA and the history of the
             melanogaster subgroup
JOURNAL     Proc. Natl. Acad. Sci. U.S.A. 87 (24), 9558-9562 (1990)
MEDLINE     91088557
AUTHORS     7 (bases 14215 to 14512)
             Ballard,J.W., Olsen,G.J., Faith,D.P., Odgers,W.A., Rowell,D.M. and
             Atkinson,P.W.
TITLE       Evidence from 12S ribosomal RNA sequences that onychophorans are
             modified arthropods
JOURNAL     Science 258 (5086), 1345-1348 (1992)
MEDLINE     93088057
AUTHORS     8 (bases 14917 to 19517)
             Lewis,D.L., Farr,C.L., Farguhar,A.L. and Kaguni,L.S.
TITLE       Sequence, organization, and evolution of the A+T region of
             Drosophila melanogaster mitochondrial DNA
JOURNAL     Mol. Biol. Evol. 11 (3), 523-538 (1994)
MEDLINE     94285822
AUTHORS     9 (bases 1 to 408; 13319 to 19517)
             Lewis,D.L., Farr,C.L. and Kaguni,L.S.
TITLE       Drosophila melanogaster mitochondrial DNA: completion of the
             nucleotide sequence and evolutionary comparisons
JOURNAL     Insect Mol. Biol. 4 (4), 263-278 (1995)
MEDLINE     96423163
AUTHORS     10 (bases 1 to 19517)
             Lewis,D.L., Farr,C.L. and Kaguni,L.S.
TITLE       Direct Submission
JOURNAL     Submitted (03-OCT-1995) Laurie S. Kaguni, Biochemistry Department,
             Michigan State University, East Lansing, MI 48824-1319, USA
FEATURES
    source
        1. 19517
            /organism="Drosophila melanogaster"
            /organelle="mitochondrion"
            /db_xref="taxon:7227"
            /note="derived from new and previously submitted
            sequences; sequence is a composite containing sequences
            obtained from different Drosophila melanogaster strains"
            1. 65
                /product="tRNA-11e"
                complement(97..165)
                /product="tRNA-Gln"
                171..239
                /product="tRNA-Phe"
                240..1265
                /codon_start=1
                /transl_table=5
        CDS
            /product="NADH dehydrogenase subunit 2"
            /protein_id="AAC47811.1"
            /db_xref="GI:1166530"
            /translation="MFNSSSKILFTITMITGLITVTSNWGLGAMGLEINLLSFPL
            LKSDNNLMSTESLAKFTQVLNASTVLLFSSILMLKNNMNNNEINSFTSMIL
            LKSGAAPFHEFPMMEGLTMNALMLMTWOKTAPMLLSYLNKYLKLLISVLSYI
            GATGGLNQLSRKTKMAFSSITNLMGLSSLSIMSESTWILFFPFSLSLTFPMFIF
            KLFHLNQLSWEPYNSKILKFTLFPNLTSLGGLEPPFGFLPKHLYIOQLTCLNDYEMIF
            IMMSTLITLFFYLIRICYSAPMNNYFENNIMKMNNSINYNNYMTTFEFSIFGLFI
            SLFFMF"
            1264..1329
                /product="tRNA-Trp"
                complement(1322..1383)
                /product="tRNA-Cys"
                complement(1403..1468)
                /product="tRNA-Tyr"
                join(1470..1472,1474..3009)
                /note="mechanism underlying reading frame shift after
                first codon uncertain"
            /codon_start=1
            /transl_table=5
            /product="cytochrome c oxidase subunit I"
            /protein_id="AAC47812.2"
            /db_xref="GI:7412849"
            /translation="MSRQMLFSTNHDIGTLYIFGAGAGWGTSLSLIRAEIHPG
            ALGDDQITNIVTAHAFIMIFPMVPMIGFGNNLVPLMLGAPMAPPPRMNNSFW
            LLEPALSLLSVSMENGAGTGVTPPLSAGIAGAGADLAIFSLAGISLIGA
            VNEITIVNNRSTGISLSDRPLFVMSVVTALILLISPLVLAGATMLTDRNLTSF
            FDPAGGDPILVOHLEFMPFGHPEVYILIPFGFMISHIISOESGKETFGSLGMFYAM
            IATGIGLFTVMAHMFYMGMDVDTFRAVFTSATMIVAPGIRKIFESMLATLHGQISYS
            PALMALGFEVLEFVYGVGLGVANSVDIILHDITYVVAHRYVLSKQAVRAIMAGF
            IHMYPLFTGLTNLNNKMLKSHFIMFTIGVNLTFEPQHLGLAGAPRYSDTPDYATWN
            IVSTIGTSLSLGILFFFTIFWESLVSORQVYIPQLNNSIEWYQNTPPAHSYSELP
            LLTN"
            3012..3077
                /product="tRNA-Leu"
                3083..3767
                /note="tAA stop codon is completed by the addition of 3' A
                residues to the mRNA"
                /codon_start=1
                /transl_except=(pos:3767,aa:TERM)
                /transl_table=5
                /product="cytochrome c oxidase subunit II"
                /protein_id="AAC47813.1"
                /db_xref="GI:1166532"
                /translation="MSTWANIQLQDSASPUMEQLIFPHDHALILVMTVLGYLWF
                LFNNNVNRELLHGLIEMITLIPALITLFTLPSRLYLLDEINSVTLKSGH
                OKWMSYEYSDPNIEPDSYMIPTNELMTGFRLLDNDNVVLPMSNPRIIVTAADVI
                HSTVVALYKVDGTGFRNLQTNFTINRGLFRFGQCEICGANHSFMPYIVIESPVNV
                FIKWISSNNS"
                3768..3838
                /product="tRNA-Lys"
                3840..3906
                /product="tRNA-asp"
                3907..4068
                /codon_start=1
                /transl_table=5
                /product="ATPase 8"
                /protein_id="AAC47814.1"
                /db_xref="GI:1166533"
                /translation="MPQMAPISWLLFIIFSITFILFCSINYSTPNPSKSNLKNI
                NLNSMNKW"
                4062..4736
                /codon_start=1
                /transl_table=5
                /product="ATPase 6"
                /protein_id="AAC47815.1"
                /db_xref="GI:1166534"
                /translation="MMTNLFYSVFDPLAIFNFSIAMLSTFLGLMLPISYIWMPSRYNI
                MNNSILLTAKKEKTLGSGHNGSTFIFISLSLFPNNFGLPEFYIFTSTSLTLL
                LSLALPLMLCFMVLGWINTHQHFALHVOGPAIILMPNVCIEFTSNIIIRGTLAVR
                LTNMTAGHLLTLTIGNTGSSMSYMLMTFLMAQIALLVLESVAMIGSYVAVAVSLT
                YSEVNV"

```

[illegible][illegible]


```

* 25191 25290: gap of 100 bp
* 25291 26110: contig of 820 bp in length
* 26111 26210: gap of 100 bp
* 26211 27023: contig of 813 bp in length
* 27024 27123: gap of 100 bp
* 27124 27914: contig of 791 bp in length
* 27915 28014: gap of 100 bp
* 28015 28806: contig of 792 bp in length
* 28807 28906: gap of 100 bp
* 28907 29718: contig of 812 bp in length
* 29719 29818: gap of 100 bp
* 29819 30609: contig of 791 bp in length
* 30610 30709: gap of 100 bp
* 30710 31489: contig of 780 bp in length
* 31490 31589: gap of 100 bp
* 31590 32396: contig of 807 bp in length
* 32397 32496: gap of 100 bp
* 32497 33283: contig of 787 bp in length
* 33284 33383: gap of 100 bp
* 33384 34154: contig of 771 bp in length
* 34155 34254: gap of 100 bp
* 34255 35100: contig of 846 bp in length
* 35101 35200: gap of 100 bp
* 35201 36031: contig of 831 bp in length
* 36032 36131: gap of 100 bp
* 36132 36954: contig of 823 bp in length
* 36955 37054: gap of 100 bp
* 37055 37959: contig of 905 bp in length
* 37960 38059: gap of 100 bp
* 38060 38854: contig of 795 bp in length
* 38855 38954: gap of 100 bp
* 38955 39764: contig of 810 bp in length
* 39765 39864: gap of 100 bp
* 39865 40647: contig of 783 bp in length
* 40648 40747: gap of 100 bp
* 40748 41559: contig of 812 bp in length
* 41560 41659: gap of 100 bp
* 41660 42448: contig of 789 bp in length
* 42449 42548: gap of 100 bp
* 42549 43343: contig of 795 bp in length
* 43344 43443: gap of 100 bp
* 43444 44234: contig of 791 bp in length
* 44235 44334: gap of 100 bp
* 44335 45143: contig of 809 bp in length
* 45144 45243: gap of 100 bp
* 45244 46041: contig of 798 bp in length
* 46042 46141: gap of 100 bp
* 46142 46937: contig of 796 bp in length
* 46938 47037: gap of 100 bp
* 47038 47830: contig of 793 bp in length
* 47831 47930: gap of 100 bp
* 47931 48734: contig of 804 bp in length
* 48735 48834: gap of 100 bp
* 48835 49633: contig of 799 bp in length
* 49634 49733: gap of 100 bp
* 49734 50517: contig of 784 bp in length
* 50518 50617: gap of 100 bp
* 50618 51470: contig of 853 bp in length
* 51471 51570: gap of 100 bp
* 51571 52367: contig of 817 bp in length
* 52368 52487: gap of 100 bp
* 52488 53299: contig of 812 bp in length
* 53300 53399: gap of 100 bp
* 53400 54204: contig of 805 bp in length
* 54205 54304: gap of 100 bp
* 54305 55129: contig of 825 bp in length
* 55130 55229: gap of 100 bp
* 55230 56063: contig of 834 bp in length
* 56064 56163: gap of 100 bp
* 56164 56939: contig of 776 bp in length
* 56940 57039: gap of 100 bp
* 57040 57851: contig of 812 bp in length
* 57852 57951: gap of 100 bp

```

```

* 57952 58782: contig of 831 bp in length
* 58783 58882: gap of 100 bp
* 58883 59712: contig of 830 bp in length
* 59713 59812: gap of 100 bp
* 59813 60604: contig of 792 bp in length.
FEATURES
    source
        1. 60604
        /organism="Homo sapiens"
        /db_xref="taxon:9606"

```

```

Query Match          9.6%; Score 57.2; DB 2; Length 60604;
Best Local Similarity 43.1%; Pred. No. 0.37;
Matches 185; Conservative 0; Mismatches 244; Indels 0; Gaps 0;

```

```

QY 134 aaacttgcttattgttgggtgacatatttcacacaaatgtaattatagtgct 193
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 10327 AATTAATATATATTTTNNATTTTNNAAATTTTNNATATATATATTTTAAATTTATTA 10386
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

QY 194 ggtgacagagatgcaatcttctgctcctaaagtcgctgcaagtaaaaaaacacc 253
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

Db 10387 TTATTTNNATTTTNTTTTNTTTTNTTTTATTTTAAATTAATTTATTTTAAATTTAAAT 10446
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

QY 254 ttctcttcaatgacgacgtagtgaggtttcttcaacttaaacacaaatgct 313
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 10447 TTTATTTNTATTTATNNAAATTTTAAATTTATTTATTTATTTATTTATTTATTTTNTT 10506
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

QY 314 aaatcatgtgtctacgttagttataattacgctataattcccaatgaatgac 373
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 10507 TAAATTTTNTTTTNTTTTNTTTTNTTTTNTTTTATTTATTTATTTTNTTTTNTTTNT 10566
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

QY 374 agaacgacatcttaatcatcatgctctgcacatgctcttcaacttaacatactctt 433
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 10567 TTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTNNANAAATTTTAAATTTATTTATTTT 10626
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

QY 434 tcgacgaatgtaaaaggaatgataaattagttatcataagtgacgctgtaaatgac 493
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 10627 TTTTATTTATTAATTTTAAATTTTATTTTATTTTATTTTATTTTATTTATTTATTTT 10686
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

QY 494 taaatctacttctgcaattaaaggcttcagacactgtgaaactttttacttttt 553
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 10687 TTTTNTTTTATTAATTTTATTTATTTTAAATTAATTAATTAATTAATTAATTTTCTTAAAT 10746
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

QY 554 tcggaataa 562
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 10747 TTATTAATA 10755

```

RESULT 14

```

AC012204 183813 bp DNA linear HTG 28-FEB-2000
LOCUS AC012204/c
DEFINITION Homo sapiens chromosome 4 clone RP11-315D13 map 4, WORKING DRAFT
SEQUENCE, 21 unordered pieces.
ACCESSION AC012204.3 GI:7107763
VERSION AC012204
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens

```

REFERENCE

```

AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE 1 (bases 1 to 183813)
JOURNAL Homo sapiens chromosome 4, clone RP11-315D13
REFERENCE Unpublished
AUTHORS 2 (bases 1 to 183813)
    Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
    Baldwin,J., Barina,N., Beckerly,R., Boguslavsky,L., Bonkhallier,B.,
    Brown,A., Castle,A., Colangelo,M., Collins,S., Collamore,A.,
    Cooke,P., Dearellano,K., Dewar,K., Domino,M., Donellan,L., Doyle,M.,
    Ferreira,P., Fitzhugh W., Forrest,C., Funke,R., Gage,D.,
    Galagan,J., Gardyna,S., Grant,G., Hagos,B., Harford,A., Horton,L.,
    Howland,J.C., Johnson,R., Jones,C., Kann,L., Kariatas,A., Klein,J.,
    Lehoczy,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
    McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J.,

```

TITLE
JOURNAL

Morrow, J., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
Stange-Rhoman, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Testaye, S., Tittel, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
Wyman, D., Ye, W.-J., Zimmer, A. and Zody, M.
Direct Submission

COMMENT

Submitted (21-OCT-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Feb 28, 2000 this sequence version replaced g1:6524238.
All repeats were identified using RepeatMasker:
Smith, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www.seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center Project name: L1740

Center Clone name: 315.D.13

Summary Statistics

Sequencing vector: M13; M7815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 136434 bases at least Q40

Consensus quality: 161213 bases at least Q30

Insert size: 190000; agarose-fp

Insert size: 181813; sum-of-coverage

Quality coverage: 3.6 in Q20 bases; agarose-fp

Quality coverage: 3.8 in Q20 bases; sum-of-coverage

* NOTE: This is a 'working draft' sequence. It currently
* consists of 21 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```
1      1240: contig of 1240 bp in length
*      1241 1340: gap of 100 bp
*      1341 2108: contig of 768 bp in length
*      2109 2208: gap of 100 bp
*      2209 3537: contig of 1329 bp in length
*      3538 3637: gap of 100 bp
*      3638 6473: contig of 2836 bp in length
*      6474 6573: gap of 100 bp
*      6574 10752: contig of 4179 bp in length
*      10753 10852: gap of 100 bp
*      10853 15130: contig of 4278 bp in length
*      15131 15230: gap of 100 bp
*      15231 17255: contig of 2025 bp in length
*      17256 17355: gap of 100 bp
*      17356 22493: contig of 5138 bp in length
*      22494 22593: gap of 100 bp
*      22594 31215: contig of 8622 bp in length
*      31216 31315: gap of 100 bp
*      31316 36161: contig of 4846 bp in length
*      36162 36261: gap of 100 bp
*      36262 43637: contig of 7376 bp in length
*      43638 43737: gap of 100 bp
*      43738 50536: contig of 6799 bp in length
*      50537 50636: gap of 100 bp
*      50637 58712: contig of 8076 bp in length
*      58713 58812: gap of 100 bp
*      58813 68091: contig of 9279 bp in length
*      68092 68191: gap of 100 bp
*      68192 79031: contig of 10840 bp in length
*      79032 79131: gap of 100 bp
*      79132 91039: contig of 11908 bp in length
*      91040 91139: gap of 100 bp
*      91140 105647: contig of 14508 bp in length
```

```
*      105648 105747: gap of 100 bp
*      105748 121791: contig of 16044 bp in length
*      121792 121891: gap of 100 bp
*      121892 139013: contig of 17122 bp in length
*      139014 139113: gap of 100 bp
*      139114 156558: contig of 17545 bp in length
*      156559 156758: gap of 100 bp
*      156759 183813: contig of 27055 bp in length.
```

FEATURES
Location/Qualifiers

source

```
1. 183813
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/map="4"
/clone="RP11-315D13"
/clone_lib="RP11 Human Male BAC"
1. 1240
/note="assembly_fragment"
1341. 2108
/note="assembly_fragment"
clone_end:T7
vector_side:right"
2209. 3537
/note="assembly_fragment"
3638. 6473
/note="assembly_fragment"
6574. 10752
/note="assembly_fragment"
10853. 15130
/note="assembly_fragment"
15231. 17255
/note="assembly_fragment"
clone_end:SP6
vector_side:right"
17356. 22493
/note="assembly_fragment"
22594. 31215
/note="assembly_fragment"
31316. 36161
/note="assembly_fragment"
36262. 43637
/note="assembly_fragment"
43738. 50536
/note="assembly_fragment"
50637. 58712
/note="assembly_fragment"
58813. 68091
/note="assembly_fragment"
68192. 79031
/note="assembly_fragment"
79132. 91039
/note="assembly_fragment"
91140. 105647
/note="assembly_fragment"
105748. 121791
/note="assembly_fragment"
121892. 139013
/note="assembly_fragment"
139114. 156558
/note="assembly_fragment"
156759. 183813
/note="assembly_fragment"
```

```
BASE COUNT 54381 a 37506 c 35430 g 54474 t 2022 others
ORIGIN
```

```
Query Match 9.6%: Score 57; DB 2; Length 183813;
Best Local Similarity 46.5%: Pred. No. 0.31;
Matches 222; Conservative 0; Mismatches 250; Indels 5; Gaps 1;
```

```
Oy 113 acacattatctactgtgaacactgcttactgtgtgtacatatccacaa 172
Db 28742 ATAAATATATTTTATATATATATATATATATATATATATATATTTT 26683
```

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 22, 2002, 06:45:08 ; Search time 463.88 Seconds
(without alignments)
2194.814 Million cell updates/sec

Title: US-09-693-205-7_COPY_12200_12792

Perfect score: 593
Sequence: 1 acatctatgtttacagct.....acctccattttatgtctt 593

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

N_Geneseq_032802.*
1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	593	100.0	1317	22	AAS29058	CDNA encoding for
2	593	100.0	12792	22	AAH20176	Human mutated spas
3	593	100.0	12793	22	AAH20174	Human spastin nucl
4	593	100.0	12793	22	AAH20178	Human mutated spas
5	593	100.0	12793	22	AAH20179	Human mutated spas
6	593	100.0	12793	22	AAH20182	Human mutated spas
7	579.2	97.7	1387	22	AAS29132	CDNA encoding for
8	56.6	9.5	11691	24	ABL34241	Human immune syste
9	55	9.3	13038	24	ABL33274	Human immune syste

10	54.8	9.2	15387	24	ABL32184	Human immune syste
11	54.6	9.2	4468	21	AAC62126	DNA encoding a cal
12	54.6	9.2	4611	21	AAC62127	DNA encoding a cal
13	54.4	9.2	13574	21	ABL33316	Human immune syste
14	52.8	8.9	17848	22	AAS45323	Chemically pretrea
15	52.8	8.9	19087	24	ABL32792	Human immune syste
16	51.6	8.7	8237	22	AAS46802	Tumour suppressor
17	51.4	8.7	3442	23	ABL26886	Drosophila melanog
18	51.4	8.7	18512	24	ABL32976	Human immune syste
19	51.2	8.6	5487	24	ABL33598	Human immune syste
20	51	8.6	51	22	AAI76962	Human immune syste
21	51	8.6	51	22	AAI76962	Human immune syste
22	51	8.6	9652	24	ABL32908	Human immune syste
23	51	8.6	11805	24	ABL33748	Human immune syste
24	50.8	8.6	15872	22	AAS46519	Tumour suppressor
25	50.6	8.5	6503	22	ABL32771	Human immune syste
26	50.6	8.5	8169	22	AAS46287	Tumour suppressor
27	50.4	8.5	6294	24	ABL33054	Human immune syste
28	50.2	8.5	5424	24	ABL32854	Human immune syste
29	50.2	8.5	6038	24	ABL32065	Human immune syste
30	50.2	8.5	6101	24	AAS61054	Human gene regulat
31	50	8.4	17419	22	AAS45393	Chemically pretrea
32	50	8.4	17419	22	ABL33295	Human immune syste
33	50	8.4	20486	24	ABL34611	Human metastasis a
34	49.8	8.4	6112	24	ABL32489	Human immune syste
35	49.8	8.4	8423	24	ABL33407	Human immune syste
36	49.6	8.4	5357	24	ABL33546	Human immune syste
37	49.4	8.3	51	22	AAI76963	Human immune syste
38	48.8	8.2	5888	24	ABL34457	Human metastasis a
39	48.6	8.2	6109	24	ABL32326	Human immune syste
40	48.6	8.2	18624	24	AAS61077	Human gene regulat
41	48.6	8.2	18624	24	ABL33702	Human immune syste
42	48.4	8.2	7359	24	ABL33862	Human immune syste
43	48.2	8.1	6126	22	AAS46573	Tumour suppressor
44	48.2	8.1	6126	24	ABL33830	Human immune syste
45	48.2	8.1	6171	24	ABL33010	Human immune syste

ALIGNMENTS

RESULT 1
AAS29058 standard; CDNA; 1317 BP.
XX
AC AAS29058;
XX
DT 21-NOV-2001 (first entry)
XX
DE CDNA encoding for human DNA-binding protein #29.
XX
XX Human; DNA-binding protein; histone; chromo domain protein;
KW chromatin organisation modifier; y-box binding protein;
KW DNA organisation; gene transcription; malignant disease;
KW autoimmune disorder; rheumatic disease; genetic abnormality;
KW infectious disease; neurological disorder; gene therapy;
KW immunomodulatory; anti-HIV; anti rheumatic; anti microbial;
cytostatic; ss.
XX
XX Homo sapiens.
OS
XX
PN WO200155162-A1.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01305.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225457.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231442.
PR 08-SEP-2000; 2000US-0231443.
PR 08-SEP-2000; 2000US-0231444.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232297.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.

PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251889.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-46557/50.

P-PSDB; AAU18182.

Nucleic acid molecules encoding human secreted chromosomal binding proteins, used in preventing, treating or ameliorating a disorder, e.g. Alzheimer's and Parkinson's diseases and cancers -

Claim 4; SEQ ID No 39; 561pp; English.

The present invention relates to the isolation of novel DNA-binding proteins (AAU18154-AAU18281), and cDNA and genomic sequences encoding

CC for these proteins, DNA-binding proteins such as histones, chromo
CC (chromatin organisation modifier) domain proteins, and Y-box binding
CC proteins may contribute to diseases resulting from aberrant DNA
CC organisation and/or gene transcription. The sequences of the invention
CC are useful in screening assays to identify antagonists and/or agonists
CC that may enhance or block activities mediated by DNA-binding proteins.
CC Blockers of DNA-binding proteins may be useful in treating disorders
CC such as malignant diseases (e.g. cancer), autoimmune disorders
CC (e.g. diabetes mellitus), rheumatic diseases (e.g. Rheumatoid
CC arthritis), genetic abnormalities (e.g. cystic fibrosis), infectious
CC diseases (e.g. HIV) and neurological disorders (e.g. Alzheimer's
CC disease). The polynucleotide sequences of the invention may also be
CC used in gene therapy. AAS29030-AAS29157 represent cDNA sequences
CC encoding for novel DNA-binding proteins.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pat_sequences.

SQ Sequence 1317 BP; 418 A; 167 C; 243 G; 489 T; 0 other;

Query Match	100.0%	Score 593;	DB 22;	Length 1317;
Best Local Similarity	100.0%	Pred. No. 7.6e-114;		
Matches 593;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY	1	acacctatggttcaagctctctggtttgtagaagatgaacaggaacatcaaatggt	60
Db	629	acacctatggttcaagctctctggtttgtagaagatgaacaggaacatcaaatggt	688
QY	61	ggcgcttcttaaccaggtgttagatgtttcttgaaacgtctggccaagcaact	120
Db	689	ggcgcttcttaaccaggtgttagatgtttcttgaaacgtctggccaagcaact	748
QY	121	tattaactgttagaacactgtgcttaatggttgggtgtacatacttccacaatgtata	180
Db	749	tattaactgttagaacactgtgcttaatggttgggtgtacatacttccacaatgtata	808
QY	181	attatatagtggtgttggaacagagtgcaactcttgggtcctaagggtgcagttaa	240
Db	809	attatatagtggtgttggaacagagtgcaactcttgggtcctaagggtgcagttaa	868
QY	241	aaaaaacaacacctttcttccaataagagatbaatgaggttttttaacttaanaac	300
Db	869	aaaaaacaacacctttcttccaataatgagatbaatgaggttttttaacttaanaac	928
QY	301	atccaaaatigttaaatactgtgtatcatagtagtttaaatatgcgctatacttc	360
Db	929	atccaaaatigttaaatactgtgtatcatagtagtttaaatatgcgctatacttc	988
QY	361	cccatgaaatgatcagaactgcataatcaatcatggtgtgtctgcgaatgctcttaactt	420
Db	989	cccatgaaatgatcagaactgcataatcaatcatggtgtgtctgcgaatgctcttaactt	1044
QY	421	aacatattctcttggcagaatgttaaaaggtgaatgaatattgattatataagtgactgg	480
Db	1049	aacatattctcttggcagaatgttaaaaggtgaatgaatattgattatataagtgactgg	1104
QY	481	ctgtaaaatgagtcaaatataacttatagcacataaagggtctacagaacaatgttgaact	540
Db	1109	ctgtaaaatgagtcaaatataacttatagcacataaagggtctacagaacaatgttgaact	1164
QY	541	ttttactcttatatgggaataaaggaatggttcgacctccacatcttaatgctt	593
Db	1169	ttttactcttatatgggaataaaggaatggttcgacctccacatcttaatgctt	1221

RESULT	2
AAH20176	
ID	AAH20176 standard; DNA; 12792 BP.

DT 09-AUG-2001 (first entry)

XX	Human mutated spastin nucleotide sequence	SEQ ID NO:7.
DE		

Human; mouse; spastin; ARSACS; chromosome 13q11; identification;
autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;
neurodegenerative disease; reduced sensory nerve conduction; diagnosis;
reduced motor nerve velocity; hypermyelination of retinal nerve fibre;
atrophy of upper cerebellar vermis; absence of Purkinje cell;
abnormal neuronal lipid storage; genetic disorder; characterisation; ds

OS	Homo sapiens.
OS	Synthetic.

FH	Key	Location/Qualifiers

Location/Qualifiers

```
/*tag= a
```

```
/*tag= a
```

XX
XX

XX
XX
06-APR-2001

XX	20-OCT-2000: 2000WA-11S29130
DE	

00W0-11529130

XX 20-OCT-1999 0911S-0160588
 DP

99115-0160588

XX
PA (UVMC-) UNIT MCSTTJ-

GTII.

PA (HOP1-) HOPITAL SAINTE-JUSTINE:
XX
XX

SAINT-JUSTINE.

XX
F1
Hudson 10, Engel C, Richter A,
XX

ERIC, RICHEL A,

DR P-PSDB; AAB97821.

1.

PT New isolated polynucleotide, encoding spastin gene, and polypeptides.

lynuclotide, encoding spastin gene, and polypeptides,

PT Charlevoix-Saguenay disease by detecting two point mutations in spastin

enay disease by detecting two point mutations in spastin

XX
XX

11

The present invention describes human and mouse spastin, and mutated human spastin (autosomal recessive spastic ataxia of Charlevoix-Saguenay (ARSACS)) gene sequences (1). The spastin gene has been mapped to chromosome 13q11. (1) have neuroprotective activities and can be used in gene therapy and as a spastin polypeptide agonists. (1), their fragments or their complements can be useful for assaying the presence of a nucleic acid molecule in a sample. (1) is useful for diagnosing or aiding in the diagnosis of an early onset neurodegenerative disease in an individual. (1) The neurodegenerative disease comprises reduced sensory nerve conduction, reduced motor nerve velocity, hypomyelination of retinal nerve fibres, atrophy of upper cerebellar vermis, absence of Purkinje cells and abnormal neuronal lipid storage. (1) can also be used to produce antisense nucleic acids, is useful as molecular weight or chromosome markers, to identify genetic disorders, as hybridisation probes or primers, as an antigen, identify and express recombinant protein for analysis, characterisation or therapeutic use, or as markers for tissues in which the corresponding protein is expressed. Diagnostic methods from the present invention can be used to identify subjects having or at risk of developing a disease or disorder associated with aberrant expression or activity of (1). The assays can be utilised to identify a subject having or at risk of developing a disorder associated with Spastin protein or spastin gene expression or activity. The present sequence encodes a mutated human spastin.

N.B. The present sequence is not given in the present specification but is derived from the human spastin nucleotide sequence (AAH20174) as stated on page 14.

Sequence 12792 BP; 4163 A; 2256 C; 2487 G; 3886 T; 0 other;

Query Match	100.0%;	Score 593;	DB 22;	Length 12792;
Best Local Similarity	100.0%;	Pred. No. 9.7e-114;		

Query match	100.0%;	Score 553;	DB 22;
Best Local Similarity	100.0%;	Pred. No. 9.7e-114;	

Matches	593;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
QY	1	acatcttatgtttacagagcttcctctgtttgatgaagatagcaacggaacaaactcaaatggt	60						
Db	12200	acatcttatgtttacagagcttcctctgtttgatgaagatagcaacggaacaaactcaaatggt	12259						
QY	61	ggcagttcttattaccaggtttagtatgtttcttgcgaaactgcttgcgaacaacatt	120						
Db	12260	ggcagttcttattaccaggtttagtatgtttcttgcgaaactgcttgcgaacaacatt	12319						
QY	121	tatttaactgttgaagaactgcttattatgtttgtgtgtataattttcccaaatgttata	180						
Db	12320	tatttaactgttgaagaactgcttattatgtttgtgtgtataattttcccaaatgttata	12379						
QY	181	atttatatgttggttggaacagatgcacatcttctgttctaaagtgctgcagtttaa	240						
Db	12380	atttatatgttggttggaacagatgcacatcttctgttctaaagtgctgcagtttaa	12439						
QY	241	aaaaaaaaaaccttcttcttccaatagcgatgtagtggaagttttttaactttaaaac	300						
Db	12440	aaaaaaaaaaccttcttcttccaatagcgatgtagtggaagttttttaactttaaaac	12499						
QY	301	atcaaaaattgttaaatcatatggttatctagttagttataataatcggctataattc	360						
Db	12500	atcaaaaattgttaaatcatatggttatctagttagttataataatcggctataattc	12559						
QY	361	cccatgaatgacagacactgacatttaatactgttctgcgcagcttcttaactt	420						
Db	12560	cccatgaatgacagacactgacatttaatactgttctgcgcagcttcttaactt	12619						
QY	421	aacataattctcttgcagaatgttaaaagtgaaatgaatgaattatataatgaatgactg	480						
Db	12620	aacataattctcttgcagaatgttaaaagtgaaatgaatgaattatataatgaatgactg	12679						
QY	481	cgttaaatgagtctataataacttattatgaatgaagggtttacagaactgttgaactt	540						
Db	12680	cgttaaatgagtctataataacttattatgaatgaagggtttacagaactgttgaactt	12739						
QY	541	ttttacttttattggaataagaatgtttgcacctccacatttatgtctt	593						
Db	12740	ttttacttttattggaataagaatgtttgcacctccacatttatgtctt	12792						

RESULT 3

AAH20174

ID AAH20174 standard; DNA: 12793 BP.

AC AAH20174;

DT 09-AUG-2001 (first entry)

DE Human spastin nucleotide sequence SEQ ID NO:1.

XX Human; mouse; spastin; ARSACS; chromosome 13q11; identification;
KW autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;
KW neurodegenerative disease; reduced sensory nerve conduction; diagnosis;
KW reduced motor nerve velocity; hypermyelination of retinal nerve fibres;
KW atrophy of upper cerebellar vermis; absence of Purkinje cell;
KW abnormal neuronal lipid storage; genetic disorder; characterisation; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 77..11566
FT /tag="a
FT /product="spastin"
XX
PN WO200129266-A2.
XX
PD 26-APR-2001.
XX
PE 20-OCT-2000; 2000WO-US29130.
XX

PR 20-OCT-1999; 99US-0160588.
XX
XX (UYMC-) UNIV MCGILL.
PA (HOB1-) HOPITAL SAINTE-JUSTINE.
XX
XX Hudson TJ, Engert J, Richter A;
PI
XX WPI, 2001-308494/32.
DR
DR P-PSDB; AAB97819.
XX

PT New isolated polynucleotide, encoding spastin gene, and polypeptides,
PT useful for diagnosing autosomal recessive spastic ataxia of
PT Charlevoix-Saguenay disease by detecting two point mutations in spastin
PT gene sequence -
XX
XX

PS Claim 1; Fig 9; 76pp; English.

CC The present invention describes human and mouse spastin, and mutated
CC human spastin (autosomal recessive spastic ataxia of Charlevoix-Saguenay
CC (ARSACS)) gene sequences (1). The spastin gene has been mapped to
CC chromosome 13q11. (1) have neuroprotective activities and can be used in
CC gene therapy and as a spastin polypeptide agonists. (1), their fragments
CC or their complements can be useful for assaying the presence of a nucleic
CC acid molecule in a sample. (1) is useful for diagnosing or aiding in the
CC diagnosis of an early onset neurodegenerative disease in an individual.
CC The neurodegenerative disease comprises reduced sensory nerve conduction,
CC reduced motor nerve velocity, hypermyelination of retinal nerve fibres,
CC atrophy of upper cerebellar vermis, absence of Purkinje cells and
CC abnormal neuronal lipid storage. (1) can also be used to produce
CC antisense nucleic acids, is useful as molecular weight or chromosome
CC markers, to identify genetic disorders, as hybridisation probes or
CC primers, as an antigen, identify and express recombinant protein for
CC analysis, characterisation or therapeutic use, or as markers for tissues
CC in which the corresponding protein is expressed. Diagnostic methods from
CC the present invention can be used to identify subjects having or at risk
CC of developing a disease or disorder associated with aberrant expression
CC or activity of (1). The assays can be utilised to identify a subject
CC having or at risk of developing a disorder associated with spastin
CC protein or spastin gene expression or activity. The present sequence
CC encodes human spastin as given in the present invention.
XX

SQ Sequence 12793 BP; 4163 A; 2256 C; 2487 G; 3887 T; 0 other;

Query Match 100.0%; Score 593; DB 22; Length 12793;

Best Local Similarity 100.0%; Pred. No. 9 7e-114;

Matches 593; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	acatcttatgtttacagagcttcctctgtttgatgaagatagcaacggaacaaactcaaatggt	60						
Db	12201	acatcttatgtttacagagcttcctctgtttgatgaagatagcaacggaacaaactcaaatggt	12260						
QY	61	ggcagttcttattaccaggtttagtatgtttcttgcgaaactgcttgcgaacaacatt	120						
Db	12261	ggcagttcttattaccaggtttagtatgtttcttgcgaaactgcttgcgaacaacatt	12320						
QY	121	tatttaactgttgaagaactgcttattatgtttgtgtgtataattttcccaaatgttata	180						
Db	12321	tatttaactgttgaagaactgcttattatgtttgtgtgtataattttcccaaatgttata	12380						
QY	181	atttatatgttggttggaacagatgcacatcttctgttctaaagtgctgcagtttaa	240						
Db	12381	atttatatgttggttggaacagatgcacatcttctgttctaaagtgctgcagtttaa	12440						
QY	241	aaaaaaaaaaccttcttcttccaatagcgatgtagtggaagtttttaactttaaaac	300						
Db	12441	aaaaaaaaaaccttcttcttccaatagcgatgtagtggaagtttttaactttaaaac	12500						
QY	301	atcaaaaattgttaaatcatatggttatctagttagttataataatcggctataattc	360						
Db	12501	atcaaaaattgttaaatcatatggttatctagttagttataataatcggctataattc	12560						
QY	361	cccatgaatgacagacactgacatttaatactgttctgcgcagcttcttaactt	420						

|||||
Db 12561 cccatgatgatcgaactgacatttaattcattgttgcgcgcacgtcttcttactt 12620
QY 421 aacatattctcttgcgaatgtaaaagtaagataaattagttatataaagtctacg 480
Db 12621 aacatattctcttgcgaatgtaaaagtaagataaattagttatataaagtctacg 12680
QY 481 ctgtaaatgtaagtaataatacttatactatgcaattaaaggcttaacagacatgttgaactt 540
Db 12681 ctgtaaatgtaagtaataatacttatactatgcaattaaaggcttaacagacatgttgaactt 12740
QY 541 ttttactttattgggaataaagaatgttgcacccctccacattttatgctt 593
Db 12741 ttttactttattgggaataaagaatgttgcacccctccacattttatgctt 12793
RESULT 4
AAH20178
ID AAH20178 standard; DNA: 12793 BP.
XX
AC AAH20178;
DT 09-AUG-2001 (first entry)
XX
DE Human mutated spastin nucleotide sequence SEQ ID NO:11.
XX
KW Human: mouse; spastin; ARSACS; chromosome 13q11; identification;
KW autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;
KW neurodegenerative disease; reduced sensory nerve conduction; diagnosis;
KW reduced motor nerve velocity; hypermyelination of retinal nerve fibres;
KW atrophy of upper cerebellar vermis; absence of Purkinje cell;
KW abnormal neuronal lipid storage; genetic disorder; characterisation; ds.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO200129266-A2.
XX
PD 26-APR-2001.
XX
PE 20-OCT-2000; 2000WO-US29130.
XX
PR 20-OCT-1999; 99US-0160588.
XX
PA (UYMC-) UNIV MCGILL.
PA (HOP1-) HOPITAL SAINT-JUSTINE.
XX
PI Hudson TJ, Engert J, Richter A;
XX
DR WPI; 2001-308494/32.
XX
PT New isolated polynucleotide, encoding spastin gene, and polypeptides,
PT useful for diagnosing autosomal recessive spastic ataxia of
PT Charlevoix-Saguenay disease by detecting two point mutations in spastin
PT gene sequence -
XX
PS Claim 1; Page -: 76pp; English.
XX
CC The present invention describes human and mouse spastin, and mutated
CC human spastin (autosomal recessive spastic ataxia of Charlevoix-Saguenay
CC (ARSACS)) gene sequences (I). The spastin gene has been mapped to
CC chromosome 13q11. (I) have neuroprotective activities and can be used in
CC gene therapy and as a spastin polypeptide agonists. (I), their fragments
CC or their complements can be useful for assaying the presence of a nucleic
CC acid molecule in a sample. (I) is useful for diagnosing or aiding in the
CC diagnosis of an early onset neurodegenerative disease in an individual.
CC The neurodegenerative disease comprises reduced sensory nerve conduction,
CC reduced motor nerve velocity, hypermyelination of retinal nerve fibres,
CC atrophy of upper cerebellar vermis, absence of Purkinje cells and
CC abnormal neuronal lipid storage. (I) can also be used to produce
CC antisense nucleic acids, is useful as molecular weight or chromosome
CC markers, to identify genetic disorders, as hybridisation probes or
CC primers, as an antigen, identify and express recombinant protein for

CC analysis, characterisation or therapeutic use, or as markers for tissues
CC in which the corresponding protein is expressed. Diagnostic methods from
CC the present invention can be used to identify subjects having or at risk
CC of developing a disease or disorder associated with aberrant expression
CC or activity of (I). The assays can be utilised to identify a subject
CC having or at risk of developing a disorder associated with spastin
CC protein or spastin gene expression or activity. The present sequence
CC represents a mutated human spastin gene from the present invention.
CC N.B. The present sequence is not given in the present specification but
CC is derived from the human spastin nucleotide sequence (AAH20174) as
CC stated on page 14.
XX
SQ Sequence 12793 BP; 4163 A; 2255 C; 2487 G; 3888 T; 0 other;
Query Match 100.0%; Score 593; DB 22; Length 12793;
Best Local Similarity 100.0%; Pred. No. 9,7e-114;
Matches 593; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 acatctatgttttacaggctccctggttgatgaaagatgcaacgaaactccaatgtg 60
Db 12201 acatctatgttttacaggctccctggttgatgaaagatgcaacgaaactccaatgtg 12260
QY 61 ggcagttcttataccagtttgatgattgttcttggaactgcttgccaagacaactt 120
Db 12261 ggcagttcttataccagtttgatgattgttcttggaactgcttgccaagacaactt 12320
QY 121 tattaactgttagaacacttgcattatgttctgtgtacataatttcacaatgttata 180
Db 12321 tattaactgttagaacacttgcattatgttctgtgtacataatttcacaatgttata 12380
QY 181 attatatagtgtgttgaacaagatgcaactcttctgttcttaaggctgcagttaa 240
Db 12381 attatatagtgtgttgaacaagatgcaactcttctgttcttaaggctgcagttaa 12440
QY 241 aaaaaaaacaccccttcttccaatgcatgcatgtagtgagtttttaaccttaaaac 300
Db 12441 aaaaaaaacaccccttcttccaatgcatgcatgtagtgagtttttaaccttaaaac 12500
QY 301 atcaaaaattgttaaaatcatgttgttactagtagtaataatcgcgttatattc 360
Db 12501 atcaaaaattgttaaaatcatgttgttactagtagtaataatcgcgttatattc 12560
QY 361 cccatgaaatgcgaactgacatttaattcatgtttgtctgcacatgcttcttactt 420
Db 12561 cccatgaaatgcgaactgacatttaattcatgtttgtctgcacatgcttcttactt 12620
QY 421 aacatattcttcttgcaaatgttaaaagtaagataaattagttatataaagtctacg 480
Db 12621 aacatattcttcttgcaaatgttaaaagtaagataaattagttatataaagtctacg 12680
QY 481 ctgtaaatgtaagtaataatacttatactatgcaattaaaggcttaacagacatgttgaactt 540
Db 12681 ctgtaaatgtaagtaataatacttatactatgcaattaaaggcttaacagacatgttgaactt 12740
QY 541 ttttactttattgggaataaagaatgttgcacccctccacattttatgctt 593
Db 12741 ttttactttattgggaataaagaatgttgcacccctccacattttatgctt 12793
RESULT 5
AAH20179
ID AAH20179 standard; DNA: 12793 BP.
XX
AC AAH20179;
DT 09-AUG-2001 (first entry)
XX
DE Human mutated spastin nucleotide sequence SEQ ID NO:12.
XX
KW Human: mouse; spastin; ARSACS; chromosome 13q11; identification;
KW autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;
KW neurodegenerative disease; reduced sensory nerve conduction; diagnosis;

KW reduced motor nerve velocity; hypermyelination of retinal nerve fibre;
 KW atrophy of upper cerebellar vermis; absence of Purkinje cell;
 KW abnormal neuronal lipid storage; genetic disorder; characterisation; ds.
 XX Homo sapiens.
 OS Synthetic.
 XX WO200129266-A2.
 XX 26-APR-2001.
 PD 20-OCT-2000; 2000WO-US29130.
 PF 20-OCT-1999; 99US-0160588.
 XX 20-OCT-1999; 99US-0160588.
 XX (UYMC-) UNIV MCGILL.
 PA (HOP-) HOPITAL SAINTE-JUSTINE.
 XX Hudson TJ, Engert J, Richter A;
 PI MPI; 2001-308494/32.
 XX
 DR New isolated polynucleotide, encoding spastin gene, and polypeptides,
 PT useful for diagnosing autosomal recessive spastic ataxia of
 PT Charlevoix-Saguenay disease by detecting two point mutations in spastin
 PT gene sequence -
 XX
 PS Claim 1: Page -: 76pp: English.
 XX
 CC The present invention describes human and mouse spastin, and mutated
 CC human spastin (autosomal recessive spastic ataxia of Charlevoix-Saguenay
 CC (ARSAKS)) gene sequences (I). The spastin gene has been mapped to
 CC chromosome 13q11. (I) have neuroprotective activities and can be used in
 CC gene therapy and as a spastin polypeptide agonists. (I), their fragments
 CC or their complements can be useful for assaying the presence of a nucleic
 CC acid molecule in a sample. (I) is useful for diagnosing or adding in the
 CC diagnosis of an early onset neurodegenerative disease in an individual.
 CC The neurodegenerative disease comprises reduced sensory nerve conduction,
 CC reduced motor nerve velocity, hypermyelination of retinal nerve fibres,
 CC atrophy of upper cerebellar vermis, absence of Purkinje cells and
 CC abnormal neuronal lipid storage. (I) can also be used to produce
 CC antisense nucleic acids, is useful as molecular weight or chromosome
 CC markers, to identify genetic disorders, as hybridisation probes or
 CC primers, as an antigen, identify and express recombinant protein for
 CC analysis, characterisation or therapeutic use, or as markers for tissues
 CC in which the corresponding protein is expressed. Diagnostic methods from
 CC the present invention can be used to identify subjects having or at risk
 CC of developing a disease or disorder associated with aberrant expression
 CC or activity of (I). The assays can be utilised to identify a subject
 CC having or at risk of developing a disorder associated with spastin
 CC protein or spastin gene expression or activity. The present sequence
 CC encodes a mutated human spastin.
 CC N.B. The present sequence is not given in the present specification but
 CC is derived from the human spastin nucleotide sequence (AAH20174) as
 CC stated on page 14.
 XX
 XX
 SQ Sequence 12793 BP; 4163 A; 2257 C; 2487 G; 3886 T; 0 other;

Query Match 100.0%; Score 593; DB 22; Length 12793;
 Best Local Similarity 100.0%; Pred. No. 9.7e-114;
 Matches 593; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 acatctatggttaagagctctgttgaagatagcaaggaactcaaatggt 60
 DB 12201 acatcttaagtttaacaggtctctgtgatgaagatagcaaggaactcaaatggt 12260
 QY 61 ggcagttcttataccagctgttagtatgtttctgaaactgctgcgaagaacatt 120
 DB 12261 ggcagttcttataccagctgttagtatgtttctgaaactgctgcgaagaacatt 12320
 QY 121 tattaactgttagaactgtgcttatgtttgtgtacatatctccacaagtata 180

DB 12321 tattaactgttagaactgtgcttatgtttgtgtacatatctccacaagtata 12380
 QY 181 attatatagttggttgagcaagatgcaactctttgttctaaagtgctcagttaa 240
 DB 12381 attatatagttggttgagcaagatgcaactctttgttctaaagtgctcagttaa 12440
 QY 241 aaaaaaacacacctttcttcaatatgcatgtagtgaggttttttaacttaaaac 300
 DB 12441 aaaaaaacacacctttcttcaatatgcatgtagtgaggttttttaacttaaaac 12500
 QY 301 atcaaaaaattgttaaaatcaatctgttatactagtagttataattatcgctataattc 360
 DB 12501 atcaaaaaattgttaaaatcaatctgttatactagtagttataattatcgctataattc 12560
 QY 361 cccatgaatgatacgaactacattcaattcattgtgttcgcacatgctcttaatt 420
 DB 12561 cccatgaatgatacgaactacattcaattcattgtgttcgcacatgctcttaatt 12620
 QY 421 aacataattcttgcagaatgtaaaaggaatgataattagttatataagtgactcg 480
 DB 12621 aacataattcttgcagaatgtaaaaggaatgataattagttatataagtgactcg 12680
 QY 481 ctgtaaatgatacgtataatacttatacgaatgaaggcttacagaacatgtgaaact 540
 DB 12681 ctgtaaatgatacgtataatacttatacgaatgaaggcttacagaacatgtgaaact 12740
 QY 541 ttttacttttattgggaataaggaatgttgacccctccactttatgctt 593
 DB 12741 ttttacttttattgggaataaggaatgttgacccctccactttatgctt 12793

RESULT 6
 AAH20182
 ID AAH20182 standard; DNA: 12793 BP.
 XX
 XX AAH20182:
 AC
 XX
 DT 09-AUG-2001 (first entry)
 XX
 DE Human mutated spastin nucleotide sequence SEQ ID NO:15.
 XX
 KW Human; mouse; spastin; ARSAKS; chromosome 13q11; identification;
 KW autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;
 KW neurodegenerative disease; reduced sensory nerve conduction; diagnosis;
 KW reduced motor nerve velocity; hypermyelination of retinal nerve fibre;
 KW atrophy of upper cerebellar vermis; absence of Purkinje cell;
 KW abnormal neuronal lipid storage; genetic disorder; characterisation; ds.
 XX
 XX Homo sapiens.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT CDS 77..11566
 FT /tag= a
 FT /product= "mutated spastin"
 XX
 XX
 PN WO200129266-A2.
 XX
 PD 26-APR-2001.
 XX
 PF 20-OCT-2000; 2000WO-US29130.
 XX
 XX 20-OCT-1999; 99US-0160588.
 XX
 XX (UYMC-) UNIV MCGILL.
 PA (HOP-) HOPITAL SAINTE-JUSTINE.
 XX Hudson TJ, Engert J, Richter A;
 PI MPI; 2001-308494/32.
 XX P-PSDB: AAB97823.
 PT New isolated polynucleotide, encoding spastin gene, and polypeptides,

PT useful for diagnosing autosomal recessive spastic ataxia of
PT Charlevoix-Saguenay disease by detecting two point mutations in spastin
PT gene sequence -

XX Claim 1: Page -: 76pp; English.

CC The present invention describes human and mouse spastin, and mutated
CC human spastin (autosomal recessive spastic ataxia of Charlevoix-Saguenay
CC (ARSA) gene sequences (I). The spastin gene has been mapped to
CC chromosome 13q11. (I) have neuroprotective activities and can be used in
CC gene therapy and as a spastin polypeptide agonists. (II), their fragments
CC or their complements can be useful for assaying the presence of a nucleic
CC acid molecule in a sample. (II) is useful for diagnosing or aiding in the
CC diagnosis of an early onset neurodegenerative disease in an individual.
CC The neurodegenerative disease comprises reduced sensory nerve conduction,
CC reduced motor nerve velocity, hypermyelination of retinal nerve fibres,
CC atrophy of upper cerebellar vermis, absence of Purkinje cells and
CC abnormal neuronal lipid storage. (I) can also be used to produce
CC antisense nucleic acids, is useful as molecular weight or chromosome
CC markers, to identify genetic disorders, as hybridisation probes or
CC primers, as an antigen, identify and express recombinant protein for
CC analysis, characterisation or therapeutic use, or as markers for tissues
CC in which the corresponding protein is expressed. Diagnostic methods from
CC the present invention can be used to identify subjects having or at risk
CC of developing a disease or disorder associated with aberrant expression
CC or activity of (I). The assays can be utilised to identify a subject
CC having or at risk of developing a disorder associated with spastin
CC protein or spastin gene expression or activity. The present sequence
CC encodes a mutated human spastin.
CC N.B. The present sequence is not given in the present specification but
CC is derived from the human spastin nucleotide sequence (AAH20174) as
CC stated on page 14.

XX Sequence 12793 BP; 4163 A; 2257 C; 2487 G; 3886 T; 0 other;

Query Match 100.0%; Score 593; DB 22; Length 12793;
Best Local Similarity 100.0%; Pred. No. 9, 7e-114;
Matches 593; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 acatctatgtttacagagctctgtttgttgaagatagcaagcaagaaactcaaaatcgt 60
DB 12201 acatctatgtttacagagctctgtttgttgaagatagcaagcaagaaactcaaaatcgt 12260
QY 61 ggcagttcttattaccaggttgaattgttctcggaaactcgtccagacaacatt 120
DB 12261 ggcagttcttattaccaggttgaattgttctcggaaactcgtccagacaacatt 12320
QY 121 tattactgttgaagaactgttattgttgtgtgtaacattttccacaagaattata 180
DB 12321 tattactgttgaagaactgttattgttgtgtgtaacattttccacaagaattata 12380
QY 181 attatatagttgtgtgaacagatgcaactctttgttctcaaaagtgtcagttaa 240
DB 12381 attatatagttgtgtgaacagatgcaactctttgttctcaaaagtgtcagttaa 12440
QY 241 aaaaaaaacaactttcttcaataltgacatgtagtgaagtttttaacttaaaac 300
DB 12441 aaaaaaaacaactttcttcaataltgacatgtagtgaagtttttaacttaaaac 12500
QY 301 atcaaaaattgttaaaaatcattgttattatagttataatatacggctataattc 360
DB 12501 atcaaaaattgttaaaaatcattgttattatagttataatatacggctataattc 12560
QY 361 cccatgaatgacagaaacttaattcaatggttgcgcacgctcttactttt 420
DB 12561 cccatgaatgacagaaacttaattcaatggttgcgcacgctcttactttt 12620
QY 421 aacatatcttcttgcgaagtgaagaaagtgaatgaataatagttatataagttactg 480
DB 12621 aacatatcttcttgcgaagtgaagaaagtgaatgaataatagttatataagttactg 12680
QY 481 ctgtaaatgtgtctaaatataactttatgcaatgaaggtctacagaacatgttgaacct 540

DB 12681 ctgtaaatgtgtctaaatataactttatgcaatgaaggtctacagaacatgttgaacct 12740

QY 541 ttctactttattgggaataaggaatgttgcacccacacatttattgctt 593

DB 12741 ttctactttattgggaataaggaatgttgcacccacacatttattgctt 12793

RESULT 7

ID AAS29132 standard; cDNA; 1387 BP.

XX AAS29132;

XX 21-NOV-2001 (first entry)

XX cDNA encoding for human DNA-binding protein #103.

KW Human: DNA-binding protein; histone; chromo domain protein;

KW chromatin organisation modifier; Y-box binding protein;

KW DNA organisation; gene transcription; malignant disease;

KW autoimmune disorder; rheumatic disease; genetic abnormality;

KW infectious disease; neurological disorder; gene therapy;

KW immunomodulatory; anti-HIV; anti rheumatic; anti microbial;

XX cytostatic; ss.

XX Homo sapiens.

XX MO200155162-A1.

PD 02-AUG-2001.

XX 17-JAN-2001; 2001WO-US01305.

XX 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.

PR 11-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217487.

PR 14-JUL-2000; 2000US-0217496.

PR 14-JUL-2000; 2000US-0218290.

PR 26-JUL-2000; 2000US-0220963.

PR 26-JUL-2000; 2000US-0220964.

PR 14-AUG-2000; 2000US-0224518.

PR 14-AUG-2000; 2000US-0224519.

PR 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225214.

PR 14-AUG-2000; 2000US-0225266.

PR 14-AUG-2000; 2000US-0225267.

PR 14-AUG-2000; 2000US-0225268.

PR 14-AUG-2000; 2000US-0225270.

PR 14-AUG-2000; 2000US-0225447.

PR 14-AUG-2000; 2000US-0225757.

PR 14-AUG-2000; 2000US-0225758.

PR 14-AUG-2000; 2000US-0225759.

PR 18-AUG-2000; 2000US-0226279.

PR 22-AUG-2000; 2000US-0226681.

PR 22-AUG-2000; 2000US-0226686.

PR 22-AUG-2000; 2000US-0227182.

PR 23-AUG-2000; 2000US-0227009.

PR 30-AUG-2000; 2000US-0228924.

PR 01-SEP-2000; 2000US-0229287.

PR 01-SEP-2000; 2000US-0229343.

PR 01-SEP-2000; 2000US-0229344.

PR 01-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 05-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 06-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 08-SEP-2000; 2000US-0232081.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.

PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249246.
 PR 17-NOV-2000; 2000US-0249247.
 PR 17-NOV-2000; 2000US-0249249.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0251989.
 PR 06-DEC-2000; 2000US-0251989.
 PR 06-DEC-2000; 2000US-0251989.
 PR 06-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 11-DEC-2000; 2000US-0251990.
 PR 05-JAN-2001; 2001US-0259678.
 PR XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 PI Rosen CA, Barash SC, Ruben SM.
 DR WPI; 2001-465557/50.
 DR P-PSDB; AAU18256.
 DR XX
 PT Nucleic acid molecules encoding human secreted chromosomal binding
 PT proteins, used in preventing, treating or ameliorating a disorder, e.g.
 PT Alzheimer's and Parkinson's diseases and cancers -
 PT XX
 PS Claim 4; SEQ ID NO 113; 561pp; English.
 PS XX
 CC The present invention relates to the isolation of novel DNA-binding
 CC proteins (AAU18154-AAU18281), and cDNA and genomic sequences encoding
 CC for these proteins. DNA-binding proteins such as histones, chromo
 CC (chromatin organisation modifier) domain proteins, and Y-box binding
 CC proteins may contribute to diseases resulting from aberrant DNA
 CC organisation and/or gene transcription. The sequences of the invention
 CC are useful in screening assays to identify antagonists and/or agonists
 CC that may enhance or block activities mediated by DNA-binding proteins.
 CC Blockers of DNA-binding proteins may be useful in treating disorders
 CC such as malignant diseases (e.g. cancer), autoimmune disorders
 CC (e.g. diabetes mellitus), rheumatic diseases (e.g. rheumatoid
 CC arthritis), genetic abnormalities (e.g. cystic fibrosis), infectious
 CC diseases (e.g. HIV) and neurological disorders (e.g. Alzheimer's
 CC disease). The polynucleotide sequences of the invention may also be
 CC used in gene therapy. AAS29030-AAS29157 represent cDNA sequences
 CC encoding for novel DNA-binding proteins.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pcr_sequences.
 CC XX
 SQ Sequence 1387 BP; 494 A; 165 C; 238 G; 482 T; 8 other;

Query Match 97.7%; Score 579.2; DB 22; Length 1387;
 Best Local Similarity 99.3%; Pred. No. 5.4e-111;
 Matches 569; Conservative 2; Mismatches 1; Indels 1; Gaps 1;

OY 1 acatctatgtttacagctctcgtttgatgaagatagcaacgaaactcaaaatggt 60
 Db 620 acatctatgtttacagctctcgtttgatgaagatagcaacgaaactcaaaatggt 679
 OY 61 ggcagttctattacagctgtttagatgtttcttgcgaacgctgtgcgaagcaacatt 120
 Db 680 ggcagttctattacagctgtttagat-tgtttcttgcgaacgctgttcgaagcaacatt 738

QY 121 tattaactgttagaacacttgcttactatgttgtgtgtacataatttccacaatgttata 180
 |||||
 Db 739 nattaactgttagaacacttgcttactatgttgtgtgtacataatttccacaatgttata 798
 QY 181 attataatgtgtgtgtgaacagatgcaactcttctgtctaaaggtgtgtcaatga 240
 |||||
 Db 799 attataatgtgtgtgtgaacagatgcaactcttctgtctaaaggtgtgtcaatga 858
 QY 241 aaaaaaacacacttcttcttcaataatgcatgtgaggttttttaacttaaaac 300
 |||||
 Db 859 aaaaaaacacacttcttcttcaataatgcatgtgaggttttttaacttaaaac 918
 QY 301 atcaaaaattgttaaatcatcttggttactagtagtcaatcaatcatcggttatattc 360
 |||||
 Db 919 atcaaaaattgttaaatcatcttggttactagtagtcaatcaatcatcggttatattc 978
 QY 361 cccatgaatgcatcagaactgacatttaactcaatgcttctgcgcagatgcttactt 420
 |||||
 Db 979 cccatgaatgcatcagaactgacatttaactcaatgcttctgcgcagatgcttactt 1038
 QY 421 aacatattcttctgtcgaatgataaaggtatgataatagttatataagtgactgg 480
 |||||
 Db 1039 aacatattcttctgtcgaatgataaaggtatgataatagttatataagtgactgg 1098
 QY 481 ctgttaattgtgtctaaataacttactatgcaattgaagggttacaagacatgttaact 540
 |||||
 Db 1099 ctgttaattgtgtctaaataacttactatgcaattgaagggttacaagacatgttgaact 1158
 QY 541 ttttacttcttattgtggaataaggaatgtgtgcacccacacatttattgctt 593
 |||||
 Db 1159 ttttacttcttattgtggaataaggaatgtgtgcacccacacatttattgctt 1211

RESULT 8

ABL34241
 ID ABL34241 standard; DNA: 11691 BP.

XX ABL34241.

XX 26-MAR-2002 (first entry)

DE Human immune system associated gene SEQ ID NO: 2214.

XX Human: immune system disease; cytosine methylation; antiasthmatic;
 KW antiarteriosclerotic; antihaemic; cytosolic; noctropic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
 KW gene; ds.

XX Homo sapiens.

OS WO200200928-A2.

PN 03-JAN-2002.

XX 02-JUL-2001; 2001WO-EP07537.

XX 30-JUN-2000; 2000DE-1032529.

PR 01-SEP-2000; 2000DE-1043826.

XX (EPIC-) EPIDENOMICS AG.

PI Olek A, Plienbrock C, Berlin K;

XX WPI, 2002-130909/17.

XX Nucleic acid comprising fragment of chemically modified gene, useful
 PT for diagnosis and treatment of diseases associated with abnormal
 PT cytosine methylation -
 XX

PS Claim 1; SEQ ID NO 2214; 32pp + Sequence Listing; German.

XX The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention.

XX Sequence 11691 BP; 3695 A; 49 C; 1861 G; 6086 T; 0 other;

Query Match 9.5%; Score 56.6; DB 24; Length 11691;

Best Local Similarly 46.0%; Pred. No. 0.0085;

Matches 233; Conservative 0; Mismatches 269; Indels 5; Gaps 1;

QY 79 ttgttagtattgttcttgcgaacactgctgcgaagacacatttactgttgaacac 138
 |||||
 Db 5647 tttcttaattctttaggaataagtgagacttttgagcttacttactttagctatattcg 5706
 QY 139 ttgttactgttctgtgtacatatcttccacaatgtcatatattatagtcgttg 198
 |||||
 Db 5707 ttttttagtattttttatagattaggtttagaagatgtatttacttaatttataa 5766
 QY 199 aacagatgcacacttctgtgtctaaagggtgtcgaatlaaaaaaacacacttcc 258
 |||||
 Db 5767 agtatatgaatatagttacttttaaaagttacttaaaagttacttaaaagatgtatt 5826
 QY 259 ttcaatatgcagtgatgtgaggttttttactttaaaacacacaaatgtttaaact 318
 |||||
 Db 5827 tggaaattttttatattttaaattatataaaaaagttatttattgttgaataat 5886
 QY 319 catgtgtatctagtagttatataatcatcgcgtatatacccatgaatgacgaac 378
 |||||
 Db 5887 tttttatataatataagtttgagatttgcgttaactatttattatttatttataa 5946
 QY 379 tgacatttaattcagtggtgtgcgcacgtcttacttcaacatatttcttggcag 438
 |||||
 Db 5947 gttattttaaagaaggtatataatataatattatattatagtttttttggct 6006
 QY 439 aatgtaaaggtatgaattagttatataaagtgatcgtgcgtgaatgagctaaat 498
 |||||
 Db 6007 tattt-----ttactgttaacttatttacttagatagagttttagggattagat 6061
 QY 499 atactttatgaatgaaggcttacagacagtgtgaaccttttacttatttggga 558
 |||||
 Db 6062 ttaatttatattatagatttgggtttgttgggttagttagattttatttattgctt 6121
 QY 559 ataaaggaatgttgcacctccacatt 585
 |||||
 Db 6122 attggaagttttagtttggaaatat 6148

RESULT 9

ABL33274
 ID ABL33274 standard; DNA: 13038 BP.

XX ABL33274.

XX 26-MAR-2002 (first entry)

DE Human immune system associated gene SEQ ID NO: 1247.

XX Human: immune system disease; cytosine methylation; antiasthmatic;
 KW antiarteriosclerotic; antihaemic; cytosolic; noctropic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
 KW gene; ds.

XX Homo sapiens.
CS
XX
PN WO200200928-A2.
XX
XX
PD 03-JAN-2002.
XX
PE 02-JUL-2001; 2001WO-EP07537.
XX
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
XX
PA (EPIC-) EPIGENOMICS AG.
XX
PL Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful
XX for diagnosis and treatment of diseases associated with abnormal
XX cytosine methylation -
PS
XX
XX Claim 1; SEQ ID NO 1247; 32pp + Sequence Listing; German.
XX
XX The present invention provides a number of human immune system associated
XX genes which are modified by the methylation of cytosines. The sequences
XX can be used in the diagnosis and treatment of immune system disorders,
XX including eye diseases such as retinopathy, neovascular glaucoma and
XX macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
XX leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
XX rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
XX diseases. The present sequence is a gene of the invention.
XX
XX Sequence 13038 BP; 4059 A; 86 C; 2457 G; 6436 T; 0 other;

Query Match	9.3%	Score 55	DB 24	Length 13038
Best Local Similarity	52.4%	Pred. NO. 0.018		
Matches 144	Conservative	0	Mismatches 130	Indels 1
OY	77 agtcttgatgattgtttcccggaactgcttcgcaagacacagttaattactgttaga-a 135			
Db	4440 attttttgtaattgaaatgaaagagtttttaattgaagatttaataatggttttagt 4499			
OY	136 caactgctttaatgtttgtgtgtacatatttcccaaaagttaataattatagtggtg 195			
Db	4500 tgaagtgggttaataatgaattatgaatttaattttgttaattagtggttaattta 4559			
OY	136 ttgaacaggaatgaacatttgttgcgtctaaaggtcgtcagttaaaaaaacaacctt 255			
Db	4560 ttaaaagatgattgtttttttttttgttattgttgaataataaaaaaatatgagtt 4619			
OY	256 ttcttcaatagcagatgagtgagtggttttttaactttaaaactcaaaatgttta 315			
Db	4620 ttttttgattttcgtgaatagttggttaatttatattgaaggagagaaatgagtgtat 4679			
OY	316 aatcatctgtatcatcagtagtttaataatcag 350			
Db	4680 aattatgtttgagatatgaattattttgttcgg 4714			
RESULT 10				
ABL32184				
ID	ABL32184 standard; DNA; 15387 BP.			
AC	ABL32184;			
XX				
DT	26-MAR-2002 (first entry)			
XX				
DE	Human immune system associated gene SEQ ID NO: 157.			
XX				
FW	Human; immune system disease; cytosine methylation; antiasthmatic; antiarteriosclerotic; antinaemic; cytosinatic; nootropic;			

KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 OS antirheumatic; antiarthritic; antidiabetic; antipsoriatric;
 PN antinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
 KW gene; ds.
 XX
 OS Homo sapiens.
 PN WO200200928-A2.
 XX
 PD 03-JAN-2002.
 XX
 PF 02-JUL-2001; 2001WO-EP07537.
 XX
 PR 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 XX
 PA (EPIG-) EPIGENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K;
 DR WPI: 2002-130909/17.
 PT Nucleic acid comprising fragment of chemically modified gene, useful
 PT for diagnosis and treatment of diseases associated with abnormal
 PT cytosine methylation -
 XX
 PS *Claim 1; SEQ ID NO 157; 32pp + Sequence Listing; German.
 XX
 CC The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention.
 XX
 SQ Sequence 15387 BP; 3742 A; 300 C; 3619 G; 7726 T; 0 other;

[illegible]

QY 465 tatataagtgactgctgtaaatgargcctaatacttatagtacgaattgaaggcttaca 524
 Db 10216 tttaattgtatcttttattatagatatgaattgttaaaatttattttttttttt 10275
 QY 525 gaacatgtgaacatttttactttattcgggaataaggaaatgt 570
 Db 10276 gtattattgttatttttttattatagatatgaattgttaaaattt 10321
 RESULT 11
 AAC62126
 ID AAC62126 standard; DNA: 4468 BP.
 XX AAC62126;
 AC
 XX 06-MAR-2001 (first entry)
 DT
 XX
 DE DNA encoding a calcium-dependent serine-protease designated Pf-SUB2.
 XX
 KW Calcium-dependent serine-protease; Pf-SUB2; merozoite differentiation;
 XX major surface protein 1; MSP1-42; erythrocyte entry; malaria; ss.
 XX Plasmodium falciparum.
 OS
 FH Key Location/Qualifiers
 FT CDS 61..4074
 FT /*tag= a
 FT /product= "serine protease"
 XX
 PN FR2791685-A1.
 PD 06-OCT-2000.
 XX
 PF 31-MAR-1999; 99FR-0004039.
 XX
 PR 31-MAR-1999; 99FR-0004039.
 XX
 PA (INSP) INST PASTEUR.
 PA (CNRS) CNRS CENT NAT RECH SCI.
 XX
 PI Barale JC, Langsley G, Braun BC, Pereira Da Silva L, Blisnick T;
 XX
 DR WPI: 2000-658021/64.
 DR P-PSDB; AAB30504.
 XX
 PT Polypeptide with calcium-dependent serine-protease activity, for the
 PT prevention, treatment, and detection of malarial infections due to
 PT Plasmodium falciparum
 XX
 PS Claim 3; Page 33-34; 47pp; French.
 XX
 CC The present sequence encodes a polypeptide (Pf-SUB2) which has a
 CC calcium-dependent serine-protease activity. The Pf-SUB2 gene is
 CC expressed during the differentiation phase of merozoites. The protein
 CC is implicated in maturation of the major surface protein 1 of the
 CC merozoites (MSP1-42). The enzyme is also crucial for entry of the
 CC parasite into erythrocytes. The polypeptides and polynucleotides are
 CC used to identify inhibitors of Pf-SUB2. These inhibitors e.g. antibodies,
 CC are used for the detection, prevention, and treatment of malaria due to
 CC Plasmodium falciparum infection.
 CC
 SQ Sequence 4468 BP; 1902 A; 456 C; 641 G; 1469 T; 0 other:

Query Match 9.2%; Score 54.6; DB 21; Length 4468;
 Best Local Similarity 50.2%; Pred. No. 0.02;
 Matches 135; Conservative 0; Mismatches 134; Indels 0; Gaps 0;

QY 164 ttccacaacatgtataattatataagtgctggtgaacagatgaacattgttgctc 223
 Db 4128 tatataataattcatattattattagtaataataataatcaccctttattttt 4187

QY 224 aaagtgctgcagttaaaacacacacccctttcttcaatagtcagtgaggatt 283
 Db 4188 aaacatttcttggtatataaatacatatataattttttaaatttattcagtcattt 4247
 QY 284 tttttaactttaaanaacataaattgtttaaaatcattgtgtatcagtagttataa 343
 Db 4248 attatataatataatataatataatataatataatataatataatataatattatgatat 4307
 QY 344 ttatcgcttataattcccatgaatgatacagaactgaacatttaattcaattgttcgcg 403
 Db 4308 ttattttttaatagtaactcatttttattgtgaaacacattacccttttttcgttt 4367
 QY 404 ccatgctcttactttaacattttctt 432
 Db 4368 ttcaattattatttattattattattt 4396
 RESULT 12
 AAC62127
 ID AAC62127 standard; DNA: 4611 BP.
 XX AAC62127;
 AC
 XX 06-MAR-2001 (first entry)
 DT
 XX
 DE DNA encoding a calcium-dependent serine-protease designated Pf-SUB2.
 XX
 KW Calcium-dependent serine-protease; Pf-SUB2; merozoite differentiation;
 XX major surface protein 1; MSP1-42; erythrocyte entry; malaria; ss.
 XX Plasmodium falciparum.
 OS
 FH Key Location/Qualifiers
 FT CDS 124..4217
 FT /*tag= a
 FT /product= "serine protease"
 FT /note= "contains an intron"
 FT exon 124..2877
 FT /*tag= b
 FT Intron 2878..2930
 FT /*tag= c
 FT exon 2931..4217
 FT /*tag= d
 XX
 PN FR2791685-A1.
 PD 06-OCT-2000.
 XX
 PF 31-MAR-1999; 99FR-0004039.
 XX
 PR 31-MAR-1999; 99FR-0004039.
 XX
 PA (INSP) INST PASTEUR.
 PA (CNRS) CNRS CENT NAT RECH SCI.
 XX
 PI Barale JC, Langsley G, Braun BC, Pereira Da Silva L, Blisnick T;
 XX
 DR WPI: 2000-658021/64.
 DR P-PSDB; AAB30505.
 XX
 PT Polypeptide with calcium-dependent serine-protease activity, for the
 PT prevention, treatment, and detection of malarial infections due to
 PT Plasmodium falciparum
 XX
 PS Claim 3; Page 34-45; 47pp; French.
 XX
 CC The present sequence encodes a polypeptide (Pf-SUB2) which has a
 CC calcium-dependent serine-protease activity. The Pf-SUB2 gene is
 CC expressed during the differentiation phase of merozoites. The protein
 CC is implicated in maturation of the major surface protein 1 of the
 CC merozoites (MSP1-42). The enzyme is also crucial for entry of the
 CC parasite into erythrocytes. The polypeptides and polynucleotides are
 CC used to identify inhibitors of Pf-SUB2. These inhibitors e.g. antibodies,

CC are used for the detection, prevention, and treatment of malaria due to
CC Plasmodium falciparum infection.

XX Sequence 4611 BP; 1951 A; 459 C; 657 G; 1544 T; 0 other;

Query Match 9.2%; Score 54.6; DB 21; Length 4611;
Best Local Similarity 50.2%; Pred. No. 0.02;
Matches 135; Conservative 0; Mismatches 134; Indels 0; Gaps 0;

QY 164 ttccacaatgtataattatagtggtgtgaacagatgcaatcttggctc 223
DB 4271 tatatatatttcatatttatttagtaataataaaccttattttt 4330
QY 224 aagggtgcagtgtaaaaaaacacaccttcttcaatgcatgagtggt 283
DB 4331 aaatatttcttgataaaaatacatatattttaaatttaagcacttc 4390
QY 284 tttaacttaaaaacatcaaatgttaaaatcattgtaacatgagttataa 343
DB 4391 attatatatatatatatatatatatatatatatatatatattgtatc 4450
QY 344 ttctcgctataattcccatgagatgacagactgcaatcaatcattgtcgc 403
DB 4451 ttattttttaaagtagtcaatttttaigtgaacaacattcccttttctgtt 4510
QY 404 ccatgctcttacttaacatttctt 432
DB 4511 ttcaatttatttatttatttattt 4539

RESULT 13

ABL33316
ID ABL33316 standard; DNA: 13574 BP.

AC ABL33316;

DT 26-MAR-2002 (first entry)

XX Human immune system associated gene SEQ ID NO: 1289.

KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antihaemtic; cytosolic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antineumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW gene; ds.

XX Homo sapiens.

OS WO200200928-A2.

PN 03-JAN-2002.

PD 02-JUL-2001; 2001WO-EP07537.

PR 30-JUN-2000; 2000DE-1032529.

PR 01-SEP-2000; 2000DE-1043826.

XX (EPIC-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;

XX WPI; 2002-130909/17.

XX Nucleic acid comprising fragment of chemically modified gene, useful
PT for diagnosis and treatment of diseases associated with abnormal
XX cytosine methylation

PS Claim 1; SEQ ID NO 1289; 32pp + Sequence Listing; German.

CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention.

XX Sequence 13574 BP; 3871 A; 89 C; 2670 G; 6944 T; 0 other;

Query Match 9.2%; Score 54.4; DB 24; Length 13574;
Best Local Similarity 48.0%; Pred. No. 0.025;
Matches 191; Conservative 0; Mismatches 201; Indels 6; Gaps 1;

QY 163 ttccacaatgtataattatagtggtgtgaacagatgcaatcttggctc 222
DB 3481 tttagaanaatgttagaattagtagtaagcgtaaagtcttgattgtaagtc 3540
QY 223 taaaggtgcagtgtaaaaaaacacaccttcttcaatgcatgagtgagtc 282
DB 3541 gtttggttttaatttaaaaaataaattaggtcttaataatctttaaaaaatctt 3600
QY 283 tttaactttaaaaacacaaaattgttaaatcatggttactagtagttata 342
DB 3601 ttagtggtttttttaagaaaaatttataaaaaatgtagaatatttaagtaatttta 3660
QY 343 attacggtataattcccatgagatgacagactgcaatcaatcattgtgtcgc 402
DB 3661 gtagtagaanaatagtagatgaaatgttggaatttcttatttgtaattgtgttg 3720
QY 403 gccatgcttcttacttaacatatcttctgc-----agatgtaaaagtagat 456
DB 3721 gttgagtgttgtaataaatatttatttatttattagatatttagatagttg 3780
QY 457 aattagttataagtgtagctgctgtaaatgataatgataattacttgcattgaag 516
DB 3781 tttaagttatggtggtatagtgatgtagatagataaatttggttatataactta 3840
QY 517 ggcctacagaacatggtgaactttttactttatt 554
DB 3841 tattttagtggtgtataaatttggattattttt 3878

RESULT 14

AAS45323
ID AAS45323 standard; DNA: 17848 BP.

AC AAS45323;

DT 18-DEC-2001 (first entry)

XX Chemically pretreated complementary DNA associated with cell cycle #14.

KW Cell cycle; human; Cpc dinucleotide; cytosine methylation; HIV; aging;
KW human immunodeficiency virus; neurodegenerative disorder; solid tumour;
KW graft-versus-host disease; glomerular disease; Lewy body disease; cancer;
KW arthritis; arteriosclerosis; anti-HIV; neuroprotective; antiarthritic;
KW immunosuppressive; antitumour; cytosolic; antiarteriosclerotic; ds;
KW PCR primer.

XX Homo sapiens.

OS WO200168911-A2.

PN 20-SEP-2001.

PD 15-MAR-2001; 2001WO-EP02945.

PR 15-MAR-2000; 2000DE-1013847.

PR 06-APR-2000; 2000DE-1019058.

PR 07-APR-2000; 2000DE-1019173.

Db 14971 ttttttattttttt-----tttttttttgatcgtttttttattacaagattttt 15020

QY 446 aaggaatgataataatagttatataaagtcgtcactggtcgttaaagtcataataacttt 505

Db 15021 gaatgaagaagaatagattttttgtgt--tttaattagaattagaatttttttg 15078

QY 506 atgcataaaggcgtctacagacaacgttggaactttttactttattgggaata 561

Db 15079 attttttttatcgtatgaagcacaatttaacattttatttgattttgtaataaaa 15134

Search completed: May 22, 2002, 06:46:05
Job time: 8175 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 22, 2002, 06:34:06 ; Search time 108.82 Seconds
(without alignments)
1338.547 Million cell updates/sec

Title: US-09-693-205-7_COPY_12200_12792

Perfect score: 593
Sequence: 1 acatctatgtttacagct.....acctcacattttgtctt 593

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:*

1: /cgn2_6/prodata/1/lna/5A_COMB.seq:*
2: /cgn2_6/prodata/1/lna/5B_COMB.seq:*
3: /cgn2_6/prodata/1/lna/6A_COMB.seq:*
4: /cgn2_6/prodata/1/lna/6B_COMB.seq:*
5: /cgn2_6/prodata/1/lna/PCRUS_COMB.seq:*
6: /cgn2_6/prodata/1/lna/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	47.8	8.1	5852	1	US-07-867-106-2
2	46.6	7.9	1850	3	US-08-617-860B-32
3	46.6	7.9	4098	2	US-08-605-106-4
4	44.6	7.5	19124	2	US-08-487-826B-13
5	43.2	7.3	5820	4	US-09-029-213B-7
6	41.4	7.0	665	2	US-08-883-795A-36
7	41.4	7.0	715	4	US-08-991-789A-264
8	41.4	7.0	715	4	US-09-062-451-264
9	41.4	7.0	1511	1	US-07-891-867B-8
10	41.4	7.0	1511	1	US-08-107-755A-8
11	41.4	7.0	1511	2	US-08-544-332-8
12	41.2	6.9	767	4	US-08-998-416-472
13	40.6	6.8	9048	3	US-08-973-273-4
14	39.2	6.6	7560	4	US-09-103-478-4
15	39.2	6.6	7560	4	US-09-193-931C-4
16	39.2	6.6	6152	4	US-08-973-462-1
17	38.2	6.4	1511	1	US-07-991-867B-8
18	38.2	6.4	1511	1	US-08-107-755A-8
19	38.2	6.4	1511	2	US-08-544-332-8
20	38.2	6.4	1667	1	US-08-485-284A-1
21	38.2	6.4	2781	3	US-08-749-522-4
22	38.2	6.4	6243	2	US-09-056-075-1
23	38.2	6.4	8654	1	US-08-920-812-6
24	38.2	6.4	8654	1	US-08-920-827-6
25	38.2	6.4	8654	1	US-08-921-177-6
26	38.2	6.4	8654	1	US-08-362-577C-6
27	38.2	6.4	8654	2	US-08-920-828-6

c	28	37.6	6.3	51952	3	US-08-947-823-1	Sequence 1, Appl1
c	29	37.4	6.3	660	1	US-07-991-867B-32	Sequence 32, Appl1
c	30	37.4	6.3	660	1	US-08-107-755A-32	Sequence 32, Appl1
c	31	37.4	6.3	660	2	US-08-544-332-32	Sequence 32, Appl1
c	32	37.4	6.3	4810	3	US-08-852-629-11	Sequence 11, Appl1
c	33	37.4	6.3	4838	3	US-08-852-629-15	Sequence 15, Appl1
c	34	37.4	6.3	5852	1	US-07-867-106-2	Sequence 2, Appl1
c	35	37.2	6.3	2297	2	US-09-001-826-14	Sequence 14, Appl1
c	36	37	6.2	10091	3	US-09-058-489-34	Sequence 34, Appl1
c	37	36.8	6.2	615	1	US-08-998-416-186	Sequence 186, App
c	38	36.8	6.2	731	4	US-08-451-405A-2	Sequence 2, Appl1
c	39	36.8	6.2	5253	3	US-08-714-918-19	Sequence 19, Appl1
c	40	36.8	6.2	5253	4	US-09-265-315-19	Sequence 19, Appl1
c	41	36.8	6.2	5253	4	US-09-265-315-19	Sequence 19, Appl1
c	42	36.8	6.2	5253	4	US-09-266-417-19	Sequence 19, Appl1
c	43	36.6	6.2	665	4	US-08-998-416-937	Sequence 937, App
c	44	36.6	6.2	6243	2	US-09-056-075-1	Sequence 1, Appl1
c	45	36.6	6.2	8700	2	US-08-392-625-16	Sequence 16, Appl1

ALIGNMENTS

RESULT 1
; Sequence 2, Application US/07867106
; Patent No. 5389526
; GENERAL INFORMATION:
; APPLICANT: Slade, Martin B
; APPLICANT: Chang, Andy C M
; APPLICANT: Williams, Keith L
; TITLE OF INVENTION: Improved Plasmid Vectors for Cellular
; TITLE OF INVENTION: Slime Moulds of the Genus Dictyostellium
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5389526ris
; STREET: One Liberty Place 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/867,106
; FILING DATE: 19920625
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: AU PJ 7187
; APPLICATION NUMBER: PCT/AU90/00530
; FILING DATE: 02-NOV-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Feeney, Joanne Longo
; REGISTRATION NUMBER: 35,134
; REFERENCE/DOCKET NUMBER: RICE-0002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-568-3100
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5852 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2378..5038
; FEATURE:
; NAME/KEY: CDS

LOCATION: 2378..5038
US-07-867-106-2

Query Match 8.1%; Score 47.8; DB 1; Length 5852;
Best Local Similarity 48.4%; Pred. No. 0.011;
Matches 133; Conservative 0; Mismatches 142; Indels 0; Gaps 0;

QY 67 tctattaccagtgtagtattggttcggaacgcgttcgcaaacacattttaa 126
DB 518 TTTTGTGGGCACTTTTTCACCTTTTTCATTTTCCACCTTTTTCACCTTTT 459
QY 127 ctgttagaacactgttctggttggtgacatacttccacaaatgtataattat 186
DB 458 CTTTGTTCCTTTTTCCTTTTTCCTTTTTCCTTTTTCCTTTTTCCTTTT 399
QY 187 atagctgttggaacagagatgcaatcttctgtctaaagtgctgcagttaaaaa 246
DB 398 TAATTTATATATATCTTCATGACCTCATTTTTCATTAATTTATGTTATTA 339
QY 247 aacacacttcttcaatagcagtagtgaggttttcaacttaaacatcaaa 306
DB 338 AAATTTGATCTTTTAACTATATCTATTTTAAATTTTAAAGTTTAAACAAAG 279
QY 307 aattgtaaaatcattgtgtatctagtagttat 341
DB 278 AATTTTTCCTTTTTCCTTTTTCCTTTTTCCTTTTTCCTTTTTCCTTTT 244

RESULT 2

US-08-617-860B-32
Sequence 32, Application US/08617860B
Patent No. 613506

GENERAL INFORMATION:

APPLICANT: Topfer, R., Bautor, J., Bothmann, H., Filasak, E.,
APPLICANT: Hvilke-Grandpierre, C., Klein, B., Martini, N.,
APPLICANT: Moller, A., Schulte, W., Voeltz, M., Walek, J.,
APPLICANT: Scheil, J.,
TITLE OF INVENTION: Promoters
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: Steinberg, Raskin & Davidson, P.C.
STREET: 1140 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA

ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/617,860B

FILING DATE: 01-MAR-1996

PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/EP94/02950

FILING DATE: 05-SEP-1994

APPLICATION NUMBER: DE P4329951.2

FILING DATE: 04-SEP-1993

INFORMATION FOR SEQ ID NO: 32:

SEQUENCE CHARACTERISTICS:

LENGTH: 1850 base pairs

TYPE: Nucleic acid

STRANDEDNESS: Double stranded

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: Cuphea lanceolata

IMMEDIATE SOURCE:

LIBRARY: genomic Lambda FIX II

CLONE: C17Eg1

FEATURE:

NAME/KEY: CAAT-Signal

LOCATION: 1428..1432

NAME/KEY: TATA-Signal

LOCATION: 1553..1556

NAME/KEY: Transcription start

LOCATION: 1585

NAME/KEY: Leguminibox

LOCATION: 1642..1657

NAME/KEY: Startcodon

LOCATION: 1797..1799

NAME/KEY: CDS

LOCATION: 1797..1850

US-08-617-860B-32

Query Match 7.9%; Score 46.6; DB 3; Length 1850;
Best Local Similarity 44.8%; Pred. No. 0.017;
Matches 178; Conservative 0; Mismatches 219; Indels 0; Gaps 0;

QY 153 tgtgtacatatttccacaaagttaattatatagtggtgtaacagatgcaatc 212
DB 182 TATTTGAATATATTTTGAATTTTAAATTTTAAATTTTAAATTTTAAATTT 241
QY 213 ttgtgtctaaagtgctgcagttcaaaaaaacacacttcttcaatgcat 272
DB 242 TTTTAAAAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT 301
QY 273 gtagtgagtttcttaacttaaaaaacacaaatgttcaaatcattgtatc 332
DB 302 AAATTTAGTTTTCCTTTTAAATTTTAAATTTTAAATTTTAAATTTTAAAT 361
QY 333 gtagttataatcgtcttatattcccatgaaatgagtcgaactgcatlaa 392
DB 362 AAATTTATTTTAAAGTTTTCCTTTTAAATTTTTCGAATTTTGAATTTTGA 421
QY 393 tgtgtcgcacatgctcttacttcaacatattcttgcagatgtaaaagtaa 452
DB 422 TTGAGATTAACCGAGAAATTTATATATATATATATATATATATATATAT 481
QY 453 tgataatgattatataagtgtaactgctgtaaatgcatgcaatataactta 512
DB 482 GTCCATTTCGGTTAAACCAACGTAAGTCTGTAACGAATGATTAACGATCT 541
QY 513 taagggtcagaacatgttgaaactttttact 549
DB 542 GAAAGTTTAAAGCAACGAAAGCTATTTATTTTATTT 578

RESULT 3

US-08-605-106-4
Sequence 4, Application US/08605106
Patent No. 5910631

GENERAL INFORMATION:

APPLICANT: Topfer, R.,
APPLICANT: Martini, N.,
APPLICANT: Scheil, J.,
TITLE OF INVENTION: MEDIUM CHAIN-SPECIFIC THIOESTERS
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Schwegman, Lundberg, Woessner & Kluth, P.A.
STREET: P.O. Box 2938
CITY: Minneapolis
STATE: MN
COUNTRY: USA
ZIP: 55402

COMPUTER READABLE FORM:

```

MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/605,106
FILING DATE: 23-SEPT-1996
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/EP94/02935
FILING DATE: 01-MAR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Woessner, Warren D
REGISTRATION NUMBER: 30,440
REFERENCE/DOCKET NUMBER: 235,001US1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 612-373-6900
TELEFAX: 612-339-3061

TELEX:
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 4098 Base pairs
TYPE: nucleic acid
STRANDEDNESS: double stranded
TOPOLOGY: linear
MOLECULE TYPE: : DNS (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Cuphea lanceolata
IMMEDIATE SOURCE:
LIBRARY: genomic Lambda FIX II
CLONE: CITEg1
FEATURE:
NAME/KEY: CDS
LOCATION: join(1797..2294, 2658..2791, 2898..3011, 31322
LOCATION: ..3303, 3391..3459, 3672..3941)
FEATURE:
NAME/KEY: Startcodon
LOCATION: 1797..1799
FEATURE:
NAME/KEY: exon II
LOCATION: 1787..2294
FEATURE:
NAME/KEY: Intron II
LOCATION: 2295..2657
FEATURE:
NAME/KEY: exon III
LOCATION: 2658..2791
FEATURE:
NAME/KEY: Intron III
LOCATION: 2792..2897
FEATURE:
NAME/KEY: exon IV
LOCATION: 2898..3011
FEATURE:
NAME/KEY: Intron IV
LOCATION: 3012..3131
FEATURE:
NAME/KEY: exon V
LOCATION: 3132..3303
FEATURE:
NAME/KEY: Intron V
LOCATION: 3304..3390
FEATURE:
NAME/KEY: exon VI
LOCATION: 3391..3459
FEATURE:
NAME/KEY: Intron VI
LOCATION: 3460..3671
FEATURE:
NAME/KEY: exon VII
LOCATION: 3672..3941

```

```

;          FEATURE:
;          NAME/KEY: Stopcodon
;          LOCATION: 3942..3944
US-08-605-106-4

```

Query Match	7.9%	Score 46.6;	DB 2;	Length 4098;
Best Local Similarity	44.8%;	Pred. No. 0.019;		
Matches 178;	Conservative	0;	Mismatches 219;	Indels 0;
			Gaps	0;

Qy	153	tgtygaactatlltcccaaaatgltataatlatatagtcygtgctgaacagbatacat	212
Db	182	TATTCGAATATTTTGGAAATTTTAAATATTTTAAATATTTTAAATTTTC	241
Qy	213	tttctgtctcaagygctgcagctlaaaaaaacaaccttcttcacatagcat	272
Db	242	TTTTAAAAAATATTTTAAATATTTATATAAATTTAGTTTTTAAATTTTAAATATTTT	301
Qy	273	gtagtggagctttttaaacttlaaaaaacatcaaaaatgltbaaaatcattggtlatcta	322
Db	302	AAAAATGATTTTTTTTATTTTAAATATTTGTGAATTTTAAATATTTTGTGTTTGA	361
Qy	333	gtagtataatlatcgcttatatlltcccaatgaatgaacgacatlaattca	392
Db	362	AAAAATATTTTAAAGTTTAAATATTTTGTGAATTTTGAATATTTTGAATATTTTTC	421
Qy	393	tgttgtctgcgaatgcttcttacttaacatatllcttgcgaatgtaaaagtaa	452
Db	422	TTGGGATTAACCGGAGATTTATATATATATATATATATATATATATATATTTTC	481
Qy	453	tgaataatagttatataatgltactgycgtglaaagtcatataacttaatgcat	512
Db	482	GTTCATTTGCGTTAAACCAACGATGTCGTAAACAGATGATTAACGATCTGTGAAT	541
Qy	513	taagagcttaacagaacatgltgaacatlltttact	549
Db	542	GAATGTTTAAAGCAACGAAGCATATATTTTAAATTTT	578

```

1 RESULT 4
2 US-08-487-826B-13/c
3 ; Sequence 13, Application US/08487826B
4 ; Patent No. 5993827
5 ; GENERAL INFORMATION:
6 ; APPLICANT: Sim, Kim L.
7 ; APPLICANT: Chitnis, Chetan
8 ; APPLICANT: Miller, Louis H.
9 ; APPLICANT: Peterson, David S.
10 ; APPLICANT: Su, Xin-zhaun
11 ; APPLICANT: Wellems, Thomas E.
12 ; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
13 ; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING
14 ; PROTEINS
15 ; NUMBER OF SEQUENCES: 45
16 ; CORRESPONDENCE ADDRESS:
17 ; ADDRESSEE: Knobbe Martens Olson & Bear
18 ; STREET: 620 Newport Center Drive 16th Floor
19 ; CITY: Newport Beach
20 ; STATE: California
21 ; COUNTRY: US
22 ; ZIP: 92660
23 ; COMPUTER READABLE FORM:
24 ; MEDIUM TYPE: Floppy disk
25 ; COMPUTER: IBM PC compatible
26 ; OPERATING SYSTEM: PC-DOS/MS-DOS
27 ; SOFTWARE: Patentln Release #1.0, Version #1.25
28 ; CURRENT APPLICATION DATA:
29 ; APPLICATION NUMBER: US/08/487,826B
30 ; FILING DATE: 10-SEP-1993
31 ; CLASSIFICATION: 435
32 ; ATTORNEY/AGENT INFORMATION:
33 ; NAME: Israelsen, Ned
34 ; REGISTRATION NUMBER: 29,655
35 ; REFERENCE/DOCKET NUMBER: NIH121.001CPI
36 ;

```


STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/062,451
FILING DATE: 04-APR-1997
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121,419C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 264:
SEQUENCE CHARACTERISTICS:
LENGTH: 715 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-062-451-264

Query Match 7.0%; Score 41.4; DB 4; Length 715;
Best Local Similarity 48.7%; Pred. No. 0.22; Mismatches 117; Indels 0; Gaps 0;

Matches 111; Conservative 0; Mismatches 117; Indels 0; Gaps 0;

QY 326 tctacagtagtataatcgcgttattccccaatgaatgacgaactgacatt 385
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 157 TTTATTTTAATGATGATTTTGGATTCATTTTATATGATTTATACACT 216
QY 386 taatcagttgtctgcgcacgtctcttacttaacatattcttgcagaatgtaa 445
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 217 TTGTATTTCTTTGTTGGTCTCTCTGTCATTTCAATGTCATCTTAACTACACACA 276
QY 446 aaggtaatgataatgattataatgtagctgacgtgtaaatgagctcaataactt 505
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 277 ATCTATTTCAATTAATATCATTAACCTTACATTAATGTAAGTAACACCATATA 336
QY 506 atgcacataaggtcaccagaacatgtgaacttttttactttat 553
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 337 TTTCATTTCTCCCTTCATTCCTATGTCATTTTCTCTTAT 384

RESULT 9
US-07-991-867B-8
Sequence 8, Application US/07991867B
Patent No. 5476781
GENERAL INFORMATION:
APPLICANT: Moyer, Richard W.
APPLICANT: Hall, Richard L.
APPLICANT: Gruidl, Michael E.
TITLE OF INVENTION: No. 5476781el Entomopoxvirus Expression System
NUMBER OF SEQUENCES: 66
CORRESPONDENCE ADDRESS:
ADDRESSEE: David R. Saliwanhik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/991,867B

FILING DATE: 12-DEC-1992
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO 92/14818
FILING DATE: 12-FEB-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/827,685
FILING DATE: 30-JAN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/657,584
FILING DATE: 19-FEB-1991
ATTORNEY/AGENT INFORMATION:
NAME: Saliwanhik, David R.
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: UF114, C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 904-375-8100
TELEFAX: 904-372-5800
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 1511 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Amsacta moorei entemopoxvirus
FEATURE:
NAME/KEY: CDS
LOCATION: complement (18..218)
FEATURE:
NAME/KEY: CDS
LOCATION: complement (234..782)
FEATURE:
NAME/KEY: CDS
LOCATION: 852..1511
US-07-991-867B-8

Query Match 7.0%; Score 41.4; DB 1; Length 1511;
Best Local Similarity 47.2%; Pred. No. 0.25; Mismatches 176; Indels 2; Gaps 1;

Matches 159; Conservative 0; Mismatches 176; Indels 2; Gaps 1;

QY 176 ttaaatattatagtggttgtaacaggaatgcacatttggctcgaagggtcgta 235
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 553 TTATCTACGATATGATTTTCATTAATTAATTTTGTATTATGTAATAATTTCTTA 612
QY 236 gttaaaaaaaacaccccttcttcaalalygcagtgtagtgagtttttaactta 295
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 613 TTTAATATATTTCCGTCATGATTTATTAATTTTATTAATAAATCTATTATCTATATTA 672
QY 296 aaaaactcaaaaattgttaaaatcattggtctcagtagttataatcoggttat 355
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 673 TGA--GTATTAATTTACACATTTTGTAGATTAATAATATATCTATTATTTTCGATCA 730
QY 356 attccccaatgaatgcagaactgacatttaattcattgttgcgcacatgctctt 415
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 731 ATTCTGTGTTTGGCCAGAAAACATAGACCAATTTTATTTGATTCACGACATTTTTTTT 790
QY 416 acttaacatattcttggcagaatgtaaaagtaagataatgattgattataagtg 475
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 791 ATTAATTTGATATATTTTTCAAAATAAATAATTAATCAATGAAAAAATAAATATATCA 850
QY 476 actggtcgttaaatgagtcgaataataactttatgcaat 512
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 851 AATGATTTACTTAATTTGATTAATTTTAATTAAT 887

RESULT 10
US-08-107-755A-8
Sequence 8, Application US/08107755A
Patent No. 5721352
GENERAL INFORMATION:

```

1  APPLICANT:  Moyer, Richard W.
2  APPLICANT:  Hall, Richard U.
3  APPLICANT:  Gruidl, Michael E.
4  TITLE OF INVENTION:  No. 5721352e1 Entomopoxvirus Expression System
5  NUMBER OF SEQUENCES:  40
6  CORRESPONDENCE ADDRESS:
7  ADDRESSEE:  David R. Saliwanchik
8  STREET:  2421 N.W. 41st Street, Suite A-1
9  CITY:  Gainesville
10 STATE:  Florida
11 COUNTRY:  U.S.A.
12 ZIP:  32606
13
14 COMPUTER READABLE FORM:
15 MEDIUM TYPE:  Floppy disk
16 COMPUTER:  IBM PC compatible
17 OPERATING SYSTEM:  PC-DOS/MS-DOS
18 SOFTWARE:  Patent In Release #1.0, Version #1.25
19 CURRENT APPLICATION DATA:
20 APPLICATION NUMBER:  US/08/107,755A
21 FILING DATE:  19-AUG-1993
22 CLASSIFICATION:  435
23 PRIOR APPLICATION DATA:
24 APPLICATION NUMBER:  US 07/827,658
25 FILING DATE:  30-JAN-1992
26 PRIOR APPLICATION DATA:
27 APPLICATION NUMBER:  US 07/657,584
28 FILING DATE:  19-FEB-1991
29 ATTORNEY/AGENT INFORMATION:
30 NAME:  Saliwanchik, David R.
31 REGISTRATION NUMBER:  31,794
32 REFERENCE/DOCKET NUMBER:  U9114.C2
33 TELECOMMUNICATION INFORMATION:
34 TELEPHONE:  (904) 375-8100
35 TELEFAX:  (904) 372-5800
36 INFORMATION FOR SEQ ID NO:  8:
37 SEQUENCE CHARACTERISTICS:
38 LENGTH:  1511 base pairs
39 TYPE:  nucleic acid
40 STRANDEDNESS:  double
41 TOPOLOGY:  unknown
42 MOLECULE TYPE:  DNA (genomic)
43 ORIGINAL SOURCE:
44 ORGANISM:  Amsacta moorei entomopoxvirus
45 FEATURE:
46 NAME/KEY:  CDS
47 LOCATION:  complement (18..218)
48 FEATURE:
49 NAME/KEY:  CDS
50 LOCATION:  complement (234..782)
51 FEATURE:
52 NAME/KEY:  CDS
53 LOCATION:  852..1511
54
55 US-08-107-755A-8

```

Query Match	7.0%	Score 41.4	DB 1	Length 1511
Best Local Similarity	47.2%	Pred. No. 0.25	Mismatches 159	Conservative 0
			Indels 176	Gaps 2
176	ttataattatata	tagt	tg	tg
177	ttataattatata	tagt	tg	tg
178	ttataattatata	tagt	tg	tg
179	ttataattatata	tagt	tg	tg
180	ttataattatata	tagt	tg	tg
181	ttataattatata	tagt	tg	tg
182	ttataattatata	tagt	tg	tg
183	ttataattatata	tagt	tg	tg
184	ttataattatata	tagt	tg	tg
185	ttataattatata	tagt	tg	tg
186	ttataattatata	tagt	tg	tg
187	ttataattatata	tagt	tg	tg
188	ttataattatata	tagt	tg	tg
189	ttataattatata	tagt	tg	tg
190	ttataattatata	tagt	tg	tg
191	ttataattatata	tagt	tg	tg
192	ttataattatata	tagt	tg	tg
193	ttataattatata	tagt	tg	tg
194	ttataattatata	tagt	tg	tg
195	ttataattatata	tagt	tg	tg
196	ttataattatata	tagt	tg	tg
197	ttataattatata	tagt	tg	tg
198	ttataattatata	tagt	tg	tg
199	ttataattatata	tagt	tg	tg
200	ttataattatata	tagt	tg	tg
201	ttataattatata	tagt	tg	tg
202	ttataattatata	tagt	tg	tg
203	ttataattatata	tagt	tg	tg
204	ttataattatata	tagt	tg	tg
205	ttataattatata	tagt	tg	tg
206	ttataattatata	tagt	tg	tg
207	ttataattatata	tagt	tg	tg
208	ttataattatata	tagt	tg	tg
209	ttataattatata	tagt	tg	tg
210	ttataattatata	tagt	tg	tg
211	ttataattatata	tagt	tg	tg
212	ttataattatata	tagt	tg	tg
213	ttataattatata	tagt	tg	tg
214	ttataattatata	tagt	tg	tg
215	ttataattatata	tagt	tg	tg
216	ttataattatata	tagt	tg	tg
217	ttataattatata	tagt	tg	tg
218	ttataattatata	tagt	tg	tg
219	ttataattatata	tagt	tg	tg
220	ttataattatata	tagt	tg	tg
221	ttataattatata	tagt	tg	tg
222	ttataattatata	tagt	tg	tg
223	ttataattatata	tagt	tg	tg
224	ttataattatata	tagt	tg	tg
225	ttataattatata	tagt	tg	tg
226	ttataattatata	tagt	tg	tg
227	ttataattatata	tagt	tg	tg
228	ttataattatata	tagt	tg	tg
229	ttataattatata	tagt	tg	tg
230	ttataattatata	tagt	tg	tg
231	ttataattatata	tagt	tg	tg
232	ttataattatata	tagt	tg	tg
233	ttataattatata	tagt	tg	tg
234	ttataattatata	tagt	tg	tg
235	ttataattatata	tagt	tg	tg
236	ttataattatata	tagt	tg	tg
237	ttataattatata	tagt	tg	tg
238	ttataattatata	tagt	tg	tg
239	ttataattatata	tagt	tg	tg
240	ttataattatata	tagt	tg	tg
241	ttataattatata	tagt	tg	tg
242	ttataattatata	tagt	tg	tg
243	ttataattatata	tagt	tg	tg
244	ttataattatata	tagt	tg	tg
245	ttataattatata	tagt	tg	tg
246	ttataattatata	tagt	tg	tg
247	ttataattatata	tagt	tg	tg
248	ttataattatata	tagt	tg	tg
249	ttataattatata	tagt	tg	tg
250	ttataattatata	tagt	tg	tg
251	ttataattatata	tagt	tg	tg
252	ttataattatata	tagt	tg	tg
253	ttataattatata	tagt	tg	tg
254	ttataattatata	tagt	tg	tg
255	ttataattatata	tagt	tg	tg

Db 731 ATCTGTTGTTTGGCAGAAAAACATAGACCAATTATTATTCTATGACATTTTTTTT 750

Oy 416 acctaacatatcttccttgcagaatcgtaaaagtcataatgatgaattggtcataaagtc 475

Db 791 ATTATTGATATATTTTTTTCACAAAAAATATATATCAAGAAAAAATAAATTATCA 850

Oy 476 actgctgtaaatgactgctaaataactttagcaat 512

Db 851 AATGATTTTACTAAATTTCTGATATATTAATTTTAAAT 887

```

1 RESULT 11
2 US-08-544-332-8
3 Sequence 8, Application US/08544352
4 Patent No. 5935777
5 GENERAL INFORMATION:
6 APPLICANT: Moyer, Richard W.
7 APPLICANT: Hall, Richard L.
8 APPLICANT: Gruidl, Michael E.
9 TITLE OF INVENTION: No. 5935777el Entomopoxvirus Expression System
10 NUMBER OF SEQUENCES: 77
11 CORRESPONDENCE ADDRESS:
12 ADDRESSEE: Gerard H. Bencen
13 STREET: 2421 N.W. 41st Street, Suite A-1
14 CITY: Gainesville
15 STATE: FL
16 COUNTRY: USA
17 ZIP: 32606
18 COMPUTER READABLE FORM:
19 MEDIUM TYPE: Floppy disk
20 COMPUTER: IBM PC compatible
21 OPERATING SYSTEM: PC-DOS/MS-DOS
22 SOFTWARE: Patentin Release #1.0, Version #1.25
23 CURRENT APPLICATION DATA:
24 APPLICATION NUMBER: US/08/544,332
25 FILING DATE:
26 CLASSIFICATION: 435
27 PRIOR APPLICATION DATA:
28 APPLICATION NUMBER: US 07/991,867
29 FILING DATE: 07-DEC-1992
30 PRIOR APPLICATION DATA:
31 APPLICATION NUMBER: US 08/107,755
32 FILING DATE: 19-AUG-1993
33 PRIOR APPLICATION DATA:
34 APPLICATION NUMBER: WO 92/14818
35 FILING DATE: 12-FEB-1992
36 PRIOR APPLICATION DATA:
37 APPLICATION NUMBER: US 07/827,685
38 FILING DATE: 30-JAN-1992
39 PRIOR APPLICATION DATA:
40 APPLICATION NUMBER: US 07/657,584
41 FILING DATE: 19-FEB-1991
42 ATTORNEY/AGENT INFORMATION:
43 NAME: Bencen, Gerard H.
44 REGISTRATION NUMBER: 35,746
45 REFERENCE/DOCKET NUMBER: UP114.C4
46 TELECOMMUNICATION INFORMATION:
47 TELEPHONE: 904-375-8100
48 TELEFAX: 904-375-5800
49 INFORMATION FOR SEQ ID NO: 8:
50 SEQUENCE CHARACTERISTICS:
51 LENGTH: 1511 base pairs
52 TYPE: nucleic acid
53 STRANDEDNESS: double
54 TOPOLOGY: unknown
55 MOLECULE TYPE: DNA (genomic)
56 ORIGINAL SOURCE:
57 ORGANISM: Amsacta moorei entemopoxvirus
58 FEATURE:
59 NAME/KEY: CDS
60 LOCATION: Complement (18..218)
61 FEATURE:
62 NAME/KEY: CDS

```



```

; FILING DATE: 01-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB96/01332
; FILING DATE: 03-JUN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9511196.9
; FILING DATE: 02-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Ms Mary J Wilson
; REGISTRATION NUMBER: 32,955
; REFERENCE/DOCKET NUMBER: 620-29
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9048 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Brassica
;
US-08-973-273-4

```

```

Query Match      6.8%; Score 40.6; DB 3; Length 9048;
Best Local Similarity 43.3%; Pred. No. 0.55;
Matches 190; Conservative 0; Mismatches 249; Indels 0; Gaps 0;

```

```

QY 116 acatttacttaacgtagaacacttcttcaatgtagtgtagtgcacataatccacaatg 175
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 818 ACTATGTTGATTTAAAAAACTCTTAATTTGATTAAGATCAATCATAT 759
QY 176 ttataattatagtgtagtgaacagatgcacatcttctgttcttaagtgtagcga 235
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 758 TTATATATCAATTAATTTAGTAAATTTGCTTATTTGTCATGTTTATTAAGGTTTTAGC 699
QY 236 gttaaaaaaaacacaccttcttccaatagtagtgtagtgcacatcttcttaactta 295
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 698 TAATCATGATTTATTTATTTTACGTAAGTCACTAATTTATTTAAAAATA 639
QY 296 aaaaactcaaaaatcttcaaatcatgtagtgcacatcttctgttcttaagtgtagcga 355
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 638 CCTTAAGATCTTATATATTTAAAAATGCAATTTTATTTAGTAATATTAATTTATAT 579
QY 356 atttcccatgtagtgcagaaactgacatcttcaatgtagtgcacatcttcttctt 415
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 578 TTATATTTAAAAATTTAGTAAATTTTCTTATTTGTCATATTTATTAAGGTTTTAGACCTTT 519
QY 416 acttaccatattctcttcgagaatgtaaaggtaagtaagttaagttaagtgtagtgc 475
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 518 TACTAGGTTATGATTTATTTATTTTACTGATGTTCCGTAATGTTGTTATTTAAAA 459
QY 476 actgagctgtaaatgtagtcaatatacttcaatgtagtgcacatcttcaaggaactcaagacatgtagt 535
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 458 CAATACCTTAATGAATTTATTAATTAAGAAATTTTATTAATCACTGAATAATTTAT 399
QY 536 aacttttttacttcttact 554
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 398 ATGTATATTTAATTAATTAAT 380

```

```

RESULT 14
US-09-103-478-4/C
; Sequence 4, Application US/09103478
; Patent No. 6235975
; GENERAL INFORMATION:
; APPLICANT: Harada, John
; APPLICANT: Lotan, Tamar

```

```

; APPLICANT: Ohto, Masa-aki
; APPLICANT: Goldberg, Robert B.
; APPLICANT: Fischer, Robert L.
; TITLE OF INVENTION: LEAFY COTYLEDON1 Genes and Their Uses
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/103,478
; FILING DATE: 24-JUN-1998
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/026,221
; FILING DATE: 19-FEB-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/804,534
; FILING DATE: 21-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmhorn, Gregory P.
; REGISTRATION NUMBER: 38,440
; REFERENCE/DOCKET NUMBER: 023070-07611US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7560 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
;
US-09-103-478-4

```

```

Query Match      6.6%; Score 39.2; DB 4; Length 7560;
Best Local Similarity 52.7%; Pred. No. 1.1;
Matches 77; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

```

```

QY 271 atgtagtgaggttttttactttaaacaacataaaatgtaaaatcttgtagtacc 330
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 665 ATTTTGTAATTTTGTAGCTTACTACATCATATTTTTCGATTTAGNGATATGNT 606
QY 331 tagtagtttaataatctgcgcttatattcccatgtagtgcagaaactgacatttaatt 390
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 605 GNAAAATTTTCAATCGGANGATATTAATTAATGCGCTCAATGAAAAAACAATAATTCCT 546
QY 391 catgttctgcgcacatgctcttca 416
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 545 GAGGTAAATTTTGATATTTCTTTA 520

```

```

RESULT 15
US-09-193-931C-4/C
; Sequence 4, Application US/09193931C
; Patent No. 6320102
; GENERAL INFORMATION:
; APPLICANT: Harada, John
; APPLICANT: Lotan, Tamar
; APPLICANT: Ohto, Masa-aki
; APPLICANT: Goldberg, Robert B.
; APPLICANT: Fischer, Robert L.
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: LEAFY COTYLEDON1 Genes and Their Uses

```

```

; FILE REFERENCE: 023070-077620
; CURRENT APPLICATION NUMBER: US/09/193,931C
; CURRENT FILING DATE: 1998-11-17
; PRIOR APPLICATION NUMBER: US 09/103,478
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: US 09/026,221
; PRIOR FILING DATE: 1998-02-19
; PRIOR APPLICATION NUMBER: US 08/804,534
; PRIOR FILING DATE: 1997-02-21
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO: 4
; LENGTH: 7560
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: 7.4 kb genomic wild-type fragment containing LEC1
; NAME/KEY: misc_feature
; LOCATION: (2430)..(5824)
; OTHER INFORMATION: corresponds to SEQ ID NO:3
; NAME/KEY: promoter
; LOCATION: (2430)..(4427)
; OTHER INFORMATION: corresponds to LEC1 promoter in SEQ ID NO:3
; NAME/KEY: CDS
; LOCATION: (4427)..(5054)
; OTHER INFORMATION: LEAFY COTYLEDON1 (LEC1)
; NAME/KEY: modified base
; LOCATION: (1)..(7560)
; OTHER INFORMATION: n = g, a, c or t
; US-09-193-931C-4

```

```

Query Match          6.6%; Score 39.2; DB 4; Length 7560;
Best Local Similarity 52.7%; Pred.No.1.1;
Matches 77; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY 271 atgtagtgagagtttttaacttaaaacatcaaaatgttaaatcatgtgtatc 330
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 665 ATTTTGTGTAATTTTTCGTTTACTACATCATATTTTGTGATATTAGNGATATGNT 606

QY 331 tagtagttcataattatcggtctatattccccaatgaatgatcagaactgacatttaatt 390
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 605 GNAATAATTTTCNATCGNANNGATATATGCGCTCAATTGAAAAAACAATAATTTCTCT 546

QY 391 catgttcgtcgccatgcctctta 416
   |||||  |||||  |||||  |||||
DB 545 GAGGTTAATTTTGTGATATTCCTTA 520

```

Search completed: May 22, 2002, 06:34:29
Job time: 7574 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 22, 2002, 06:46:05 ; Search time 463.88 Seconds
(without alignments)
74.024 Million cell updates/sec

Title: US-09-693-205-47

Perfect score: 20
Sequence: 1 gtgaatgcacattgcact 20

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_032802.*
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20	100.0	20	AAH20154	Human spastin ORF
2	20	100.0	12792	AAH20176	Human spastin ORF
3	20	100.0	12793	AAH20174	Human spastin ORF
4	20	100.0	12793	AAH20178	Human spastin ORF
5	20	100.0	12793	AAH20179	Human spastin ORF
6	20	100.0	12793	AAH20182	Human spastin ORF
7	18.4	92.0	11453	AAH20175	Human spastin ORF
8	17.4	87.0	901	ABA08683	Human spastin ORF
9	16.4	82.0	1986	AAH20175	Human spastin ORF

Result No.	Score	Query Match	Length	DB ID	Description
10	16.4	82.0	2206	AAH20154	Human spastin ORF
11	16.4	82.0	2206	AAH20154	Human spastin ORF
12	16.4	82.0	2206	AAH20154	Human spastin ORF
13	16.4	82.0	2206	AAH20154	Human spastin ORF
14	16.4	82.0	2206	AAH20154	Human spastin ORF
15	16.4	82.0	2206	AAH20154	Human spastin ORF
16	16.4	82.0	2206	AAH20154	Human spastin ORF
17	16.4	82.0	2206	AAH20154	Human spastin ORF
18	16.4	82.0	2206	AAH20154	Human spastin ORF
19	16.4	82.0	2206	AAH20154	Human spastin ORF
20	16.4	82.0	2206	AAH20154	Human spastin ORF
21	15.8	79.0	336	AAH20154	Human spastin ORF
22	15.8	79.0	336	AAH20154	Human spastin ORF
23	15.8	79.0	336	AAH20154	Human spastin ORF
24	15.8	79.0	336	AAH20154	Human spastin ORF
25	15.8	79.0	336	AAH20154	Human spastin ORF
26	15.8	79.0	336	AAH20154	Human spastin ORF
27	15.4	77.0	590	AAH20154	Human spastin ORF
28	15.4	77.0	590	AAH20154	Human spastin ORF
29	15.4	77.0	590	AAH20154	Human spastin ORF
30	15.4	77.0	590	AAH20154	Human spastin ORF
31	15.4	77.0	590	AAH20154	Human spastin ORF
32	15.4	77.0	590	AAH20154	Human spastin ORF
33	15.4	77.0	590	AAH20154	Human spastin ORF
34	15.2	76.0	413	AAH20154	Human spastin ORF
35	15.2	76.0	413	AAH20154	Human spastin ORF
36	15.2	76.0	413	AAH20154	Human spastin ORF
37	15.2	76.0	413	AAH20154	Human spastin ORF
38	15.2	76.0	413	AAH20154	Human spastin ORF
39	15.2	76.0	413	AAH20154	Human spastin ORF
40	15.2	76.0	413	AAH20154	Human spastin ORF
41	15.2	76.0	413	AAH20154	Human spastin ORF
42	15.2	76.0	413	AAH20154	Human spastin ORF
43	15.2	76.0	413	AAH20154	Human spastin ORF
44	15.2	76.0	413	AAH20154	Human spastin ORF
45	15.2	76.0	413	AAH20154	Human spastin ORF

ALIGNMENTS

RESULT 1
ID AAH20154 standard; DNA: 20 BP.
AAH20154:
09-AUG-2001 (first entry)
Human spastin ORF PCR primer SEQ ID NO:47.
Human; mouse; spastin; ARSACS; chromosome 13q11; identification;
autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;
neurodegenerative disease; reduced sensory nerve conduction; diagnosis;
reduced motor nerve velocity; hypermyelination of retinal nerve fibre;
atrophy of upper cerebellar vermis; absence of Purkinje cell;
abnormal neuronal lipid storage; genetic disorder; characterisation;
PCR primer: ss.
Homo sapiens.
Synthetic.
WO200129266-A2.
26-APR-2001.
20-OCT-2000; 2000WO-US29130.
20-OCT-1999; 99US-0160588.
(UYMC-) UNIV MCGILL.
(HOPI-) HOPITAL SAINTE-JUSTINE.

PI Hudson TJ, Engert J, Richter A;
 XX WPI; 2001-308494/32.
 XX
 DR
 XX
 PT New isolated polynucleotide, encoding spastin gene, and polypeptides,
 PT useful for diagnosing autosomal recessive spastic ataxia of
 PT Charlevoix-Saguenay disease by detecting two point mutations in spastin
 PT gene sequence -
 XX
 PS Claim 23: Fig 7; 76pp; English.
 XX
 CC The present invention describes human and mouse spastin, and mutated
 CC human spastin (autosomal recessive spastic ataxia of Charlevoix-Saguenay
 CC (ARSACS)) gene sequences (I). The spastin gene has been mapped to
 CC chromosome 13q11. (I) have neuroprotective activities and can be used in
 CC gene therapy and as a spastin polypeptide agonists. (II), their fragments
 CC or their complements can be useful for assaying the presence of a nucleic
 CC acid molecule in a sample. (I) is useful for diagnosing or aiding in the
 CC diagnosis of an early onset neurodegenerative disease in an individual.
 CC The neurodegenerative disease comprises reduced sensory nerve conduction,
 CC reduced motor nerve velocity, hypermyelination of retinal nerve fibres,
 CC atrophy of upper cerebellar vermis, absence of Purkinje cells and
 CC abnormal neuronal lipid storage. (I) can also be used to produce
 CC antisense nucleic acids, is useful as molecular weight or chromosome
 CC markers, to identify genetic disorders, as hybridisation probes or
 CC primers, as an antigen, identify and express recombinant protein for
 CC analysis, characterisation or therapeutic use, or as markers for tissues
 CC in which the corresponding protein is expressed. Diagnostic methods from
 CC the present invention can be used to identify subjects having or at risk
 CC of developing a disease or disorder associated with aberrant expression
 CC or activity of (I). The assays can be utilised to identify a subject
 CC having or at risk of developing a disorder associated with Spastin
 CC protein or spastin gene expression or activity. AAH20122 to AAH20173
 CC represent specifically claimed primers which can be used in diagnostic
 CC methods from the present invention.
 XX
 SQ Sequence 20 BP; 4 A; 5 C; 5 G; 6 T; 0 other;
 XX
 Query Match 100.0%; Score 20; DB 22; Length 20;
 Best Local Similarity 100.0%; Pred. No. 0.3; Mismatches 0; Gaps 0;
 Matches 20; Conservative 0; Indels 0; Gaps 0;
 QY 1 gtgaatggccacttgctcact 20
 ||||||||||||||||
 Db 1 gtgaatggccacttgctcact 20
 XX
 RESULT 2
 AAH20176
 ID AAH20176 standard; DNA; 12792 BP.
 XX
 AC AAH20176;
 XX
 DT 09-AUG-2001 (first entry)
 XX
 DE Human mutated spastin nucleotide sequence SEQ ID NO:7.
 XX
 KW Human: mouse; spastin; ARSACS; chromosome 13q11; identification;
 KW autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;
 KW neurodegenerative disease; reduced sensory nerve conduction; diagnosis;
 KW reduced motor nerve velocity; hypermyelination of retinal nerve fibre;
 KW atrophy of upper cerebellar vermis; absence of Purkinje cell;
 KW abnormal neuronal lipid storage; genetic disorder; characterisation; ds.
 XX
 OS Homo sapiens.
 XX
 AC Synthetic.
 XX
 FH Key Location/Qualifiers
 FT 77..6604
 FT /*tag= a
 FT /product= "mutated spastin"
 XX

PN W0200129266-A2.
 XX
 XX 26-APR-2001.
 XX
 PF 20-OCT-2000; 2000WO-US29130.
 XX
 PR 20-OCT-1999; 99US-0160588.
 XX
 PA (UYMC-) UNIV MCGILL.
 PA (HOPIT-) HOPITAL SAINTE-JUSTINE.
 XX
 PI Hudson TJ, Engert J, Richter A;
 XX WPI; 2001-308494/32.
 DR P-PSDB; AAB97821.
 XX
 PT New isolated polynucleotide, encoding spastin gene, and polypeptides,
 PT useful for diagnosing autosomal recessive spastic ataxia of
 PT Charlevoix-Saguenay disease by detecting two point mutations in spastin
 PT gene sequence -
 XX
 PS Claim 1; Page -: 76pp; English.
 XX
 CC The present invention describes human and mouse spastin, and mutated
 CC human spastin (autosomal recessive spastic ataxia of Charlevoix-Saguenay
 CC (ARSACS)) gene sequences (I). The spastin gene has been mapped to
 CC chromosome 13q11. (I) have neuroprotective activities and can be used in
 CC gene therapy and as a spastin polypeptide agonists. (II), their fragments
 CC or their complements can be useful for assaying the presence of a nucleic
 CC acid molecule in a sample. (I) is useful for diagnosing or aiding in the
 CC diagnosis of an early onset neurodegenerative disease in an individual.
 CC The neurodegenerative disease comprises reduced sensory nerve conduction,
 CC reduced motor nerve velocity, hypermyelination of retinal nerve fibres,
 CC atrophy of upper cerebellar vermis, absence of Purkinje cells and
 CC abnormal neuronal lipid storage. (I) can also be used to produce
 CC antisense nucleic acids, is useful as molecular weight or chromosome
 CC markers, to identify genetic disorders, as hybridisation probes or
 CC primers, as an antigen, identify and express recombinant protein for
 CC analysis, characterisation or therapeutic use, or as markers for tissues
 CC in which the corresponding protein is expressed. Diagnostic methods from
 CC the present invention can be used to identify subjects having or at risk
 CC of developing a disease or disorder associated with aberrant expression
 CC or activity of (I). The assays can be utilised to identify a subject
 CC having or at risk of developing a disorder associated with Spastin
 CC protein or spastin gene expression or activity. The present sequence
 CC encodes a mutated human spastin.
 CC N.B. The present sequence is not given in the present specification but
 CC is derived from the human spastin nucleotide sequence (AAH20174) as
 CC stated on page 14.
 XX
 SQ Sequence 12792 BP; 4163 A; 2256 C; 2487 G; 3886 T; 0 other;
 XX
 Query Match 100.0%; Score 20; DB 22; Length 12792;
 Best Local Similarity 100.0%; Pred. No. 0.95; Mismatches 0; Gaps 0;
 Matches 20; Conservative 0; Indels 0; Gaps 0;
 QY 1 gtgaatggccacttgctcact 20
 ||||||||||||||||
 Db 6473 gtgaatggccacttgctcact 6492
 XX
 RESULT 3
 AAH20174
 ID AAH20174 standard; DNA; 12793 BP.
 XX
 AC AAH20174;
 XX
 DT 09-AUG-2001 (first entry)
 XX
 DE Human spastin nucleotide sequence SEQ ID NO:1.
 XX
 KW Human: mouse; spastin; ARSACS; chromosome 13q11; identification;
 KW

KW autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;
 KW neurodegenerative disease; reduced sensory nerve conduction; diagnosis;
 KW reduced motor nerve velocity; hypermyelination of retinal nerve fibres;
 KW atrophy of upper cerebellar vermis; absence of Purkinje cell;
 KW abnormal neuronal lipid storage; genetic disorder; characterisation; ds.
 OS Homo sapiens.

Key Location/Qualifiers
 CDS 77..11566
 /tag_a
 /product= "spastin"

MO200129266-A2.

26-APR-2001.

20-OCT-2000; 2000MO-US29130.

20-OCT-1999; 99US-0160588.

(UYMC-) UNIV MCGILL.
 (HOP1-) HOPITAL SAINTE-JUSTINE.

Hudson TJ, Engert J, Richter A;

WPI: 2001-308494/32.
 P-PSDB: AAB97819.

New isolated polynucleotide, encoding spastin gene, and polypeptides,
 useful for diagnosing autosomal recessive spastic ataxia of
 Charlevoix-Saguenay disease by detecting two point mutations in spastin
 gene sequence -

Claim 1: Fig 9; 76pp; English.

The present invention describes human and mouse spastin, and mutated
 human spastin (autosomal recessive spastic ataxia of Charlevoix-Saguenay
 (ARSACS)) gene sequences (I). The spastin gene has been mapped to
 chromosome 13q11. (I) have neuroprotective activities and can be used in
 gene therapy and as a spastin polypeptide agonists. (I), their fragments
 or their complements can be useful for assaying the presence of a nucleic
 acid molecule in a sample. (I) is useful for diagnosing or aiding in the
 diagnosis of an early onset neurodegenerative disease in an individual.
 The neurodegenerative disease comprises reduced sensory nerve conduction,
 reduced motor nerve velocity, hypermyelination of retinal nerve fibres,
 atrophy of upper cerebellar vermis, absence of Purkinje cells and
 abnormal neuronal lipid storage. (I) can also be used to produce
 antisense nucleic acids, is useful as molecular weight or chromosome
 markers, to identify genetic disorders, as hybridisation probes or
 primers, as an antigen, identify and express recombinant protein for
 analysis, characterisation or therapeutic use, or as markers for tissues
 in which the corresponding protein is expressed. Diagnostic methods from
 the present invention can be used to identify subjects having or at risk
 of developing a disease or disorder associated with aberrant expression
 or activity of (I). The assays can be utilised to identify a subject
 having or at risk of developing a disorder associated with Spastin
 protein or spastin gene expression or activity. The present sequence
 encodes human spastin as given in the present invention.

Sequence 12793 BP; 4163 A; 2256 C; 2487 G; 3887 T; 0 other;

Query Match 100.0%; Score 20; DB 22; Length 12793;
 Best Local Similarity 100.0%; Pred. No. 0.95;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gtgaatgcgcacttgcact 20
 ||||||||||||||||
 DB 6473 gtgaatgcgcacttgcact 6492

RESULT 4

AAH20178
 ID AAH20178 standard; DNA: 12793 BP.
 XX
 AC AAH20178;
 XX
 DT 09-AUG-2001 (first entry)
 XX

Human mutated spastin nucleotide sequence SEQ ID NO:11.

Human; mouse; spastin; ARSACS; chromosome 13q11; identification;
 KW autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;
 KW neurodegenerative disease; reduced sensory nerve conduction; diagnosis;
 KW reduced motor nerve velocity; hypermyelination of retinal nerve fibres;
 KW atrophy of upper cerebellar vermis; absence of Purkinje cell;
 KW abnormal neuronal lipid storage; genetic disorder; characterisation; ds.
 OS Homo sapiens.
 OS Synthetic.

MO200129266-A2.

26-APR-2001.

20-OCT-2000; 2000MO-US29130.

20-OCT-1999; 99US-0160588.

(UYMC-) UNIV MCGILL.
 (HOP1-) HOPITAL SAINTE-JUSTINE.

Hudson TJ, Engert J, Richter A;

WPI: 2001-308494/32.

New isolated polynucleotide, encoding spastin gene, and polypeptides,
 useful for diagnosing autosomal recessive spastic ataxia of
 Charlevoix-Saguenay disease by detecting two point mutations in spastin
 gene sequence -

Claim 1: Page -: 76pp; English.

The present invention describes human and mouse spastin, and mutated
 human spastin (autosomal recessive spastic ataxia of Charlevoix-Saguenay
 (ARSACS)) gene sequences (I). The spastin gene has been mapped to
 chromosome 13q11. (I) have neuroprotective activities and can be used in
 gene therapy and as a spastin polypeptide agonists. (I), their fragments
 or their complements can be useful for assaying the presence of a nucleic
 acid molecule in a sample. (I) is useful for diagnosing or aiding in the
 diagnosis of an early onset neurodegenerative disease in an individual.
 The neurodegenerative disease comprises reduced sensory nerve conduction,
 reduced motor nerve velocity, hypermyelination of retinal nerve fibres,
 atrophy of upper cerebellar vermis, absence of Purkinje cells and
 abnormal neuronal lipid storage. (I) can also be used to produce
 antisense nucleic acids, is useful as molecular weight or chromosome
 markers, to identify genetic disorders, as hybridisation probes or
 primers, as an antigen, identify and express recombinant protein for
 analysis, characterisation or therapeutic use, or as markers for tissues
 in which the corresponding protein is expressed. Diagnostic methods from
 the present invention can be used to identify subjects having or at risk
 of developing a disease or disorder associated with aberrant expression
 or activity of (I). The assays can be utilised to identify a subject
 having or at risk of developing a disorder associated with Spastin
 protein or spastin gene expression or activity. The present sequence
 represents a mutated human spastin gene from the present invention.
 N.B. The present sequence is not given in the present specification but
 is derived from the human spastin nucleotide sequence (AAH20174) as
 stated on page 14.

Sequence 12793 BP; 4163 A; 2255 C; 2487 G; 3888 T; 0 other;

Query Match 100.0%; Score 20; DB 22; Length 12793;
 Best Local Similarity 100.0%; Pred. No. 0.95;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 gtgaatgcccacttgact 20
 |||
 Db 6473 gtgaatgcccacttgact 6492

RESULT 5
 AAH20179
 ID AAH20179 standard; DNA; 12793 BP.

XX
 AC AAH20179;

DT 09-AUG-2001 (first entry)

DE Human mutated spastin nucleotide sequence SEQ ID NO:12.

XX Human; mouse; spastin; ARSACS; chromosome 13q11; identification;
 KW autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;
 KW neurodegenerative disease; reduced sensory nerve conduction; diagnosis;
 KW reduced motor nerve velocity; hypermyelination of retinal nerve fibre;
 KW atrophy of upper cerebellar vermis; absence of Purkinje cell;
 KW abnormal neuronal lipid storage; genetic disorder; characterisation; ds.

XX Homo sapiens.
 OS Synthetic.

XX WO200129266-A2.

XX 26-APR-2001.

XX 20-OCT-2000; 2000MO-US29130.

XX 20-OCT-1999; 99US-0160588.

XX (UYMC-) UNIV MCGILL.
 PA (HOP-) HOPITAL SAINTE-JUSTINE.

XX Hudson TJ, Engert J, Richter A;

XX WPI; 2001-308494/32.

XX New isolated polynucleotide, encoding spastin gene, and polypeptides,
 PT useful for diagnosing autosomal recessive spastic ataxia of
 PT Charlevoix-Saguenay disease by detecting two point mutations in spastin
 PT gene sequence -

XX Claim 1: Page -: 76pp; English.

XX The present invention describes human and mouse spastin, and mutated
 CC human spastin (autosomal recessive spastic ataxia of Charlevoix-Saguenay
 CC (ARSACS)) gene sequences (I). The spastin gene has been mapped to
 CC chromosome 13q11. (I) have neuroprotective activities and can be used in
 CC gene therapy and as a spastin polypeptide agonists. (I), their fragments
 CC or their complements can be useful for assaying the presence of a nucleic
 CC acid molecule in a sample. (I) is useful for diagnosing or aiding in the
 CC diagnosis of an early onset neurodegenerative disease in an individual.
 CC The neurodegenerative disease comprises reduced sensory nerve conduction,
 CC reduced motor nerve velocity, hypermyelination of retinal nerve fibres,
 CC atrophy of upper cerebellar vermis, absence of Purkinje cells and
 CC abnormal neuronal lipid storage. (I) can also be used to produce
 CC antisense nucleic acids, is useful as molecular weight or chromosome
 CC markers, to identify genetic disorders, as hybridisation probes or
 CC primers, as an antigen, identify and express recombinant protein for
 CC analysis, characterisation or therapeutic use, or as markers for tissues
 CC in which the corresponding protein is expressed. Diagnostic methods from
 CC the present invention can be used to identify subjects having or at risk
 CC of developing a disease or disorder associated with aberrant expression
 CC or activity of (I). The assays can be utilised to identify a subject
 CC having or at risk of developing a disorder associated with spastin
 CC protein or spastin gene expression or activity. The present sequence
 CC encodes a mutated human spastin.
 CC N.B. The present sequence is not given in the present specification but

CC is derived from the human spastin nucleotide sequence (AAH20174) as
 CC stated on page 14.
 XX
 SO Sequence 12793 BP; 4163 A; 2257 C; 2487 G; 3886 T; 0 other;

Query Match 100.0%; Score 20; DB 22; Length 12793;
 Best Local Similarity 100.0%; Pred. No. 0.95;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtgaatgcccacttgact 20
 |||
 Db 6473 gtgaatgcccacttgact 6492

RESULT 6

ID AAH20182 standard; DNA; 12793 BP.

XX
 AC AAH20182;

DT 09-AUG-2001 (first entry)

DE Human mutated spastin nucleotide sequence SEQ ID NO:15.

XX Human; mouse; spastin; ARSACS; chromosome 13q11; identification;
 KW autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;
 KW neurodegenerative disease; reduced sensory nerve conduction; diagnosis;
 KW reduced motor nerve velocity; hypermyelination of retinal nerve fibre;
 KW atrophy of upper cerebellar vermis; absence of Purkinje cell;
 KW abnormal neuronal lipid storage; genetic disorder; characterisation; ds.

XX Homo sapiens.
 OS Synthetic.

XX Key Location/Qualifiers
 FT 77..11566
 FT CDS /*tag= a
 FT /*product= "mutated spastin"

XX WO200129266-A2.

XX 26-APR-2001.

XX 20-OCT-2000; 2000MO-US29130.

XX 20-OCT-1999; 99US-0160588.

XX (UYMC-) UNIV MCGILL.
 PA (HOP-) HOPITAL SAINTE-JUSTINE.

XX Hudson TJ, Engert J, Richter A;

XX WPI; 2001-308494/32.

XX P-PDB: AAB97823.

XX New isolated polynucleotide, encoding spastin gene, and polypeptides,
 PT useful for diagnosing autosomal recessive spastic ataxia of
 PT Charlevoix-Saguenay disease by detecting two point mutations in spastin
 PT gene sequence -

XX Claim 1: Page -: 76pp; English.

XX The present invention describes human and mouse spastin, and mutated
 CC human spastin (autosomal recessive spastic ataxia of Charlevoix-Saguenay
 CC (ARSACS)) gene sequences (I). The spastin gene has been mapped to
 CC chromosome 13q11. (I) have neuroprotective activities and can be used in
 CC gene therapy and as a spastin polypeptide agonists. (I), their fragments
 CC or their complements can be useful for assaying the presence of a nucleic
 CC acid molecule in a sample. (I) is useful for diagnosing or aiding in the
 CC diagnosis of an early onset neurodegenerative disease in an individual.
 CC The neurodegenerative disease comprises reduced sensory nerve conduction,
 CC reduced motor nerve velocity, hypermyelination of retinal nerve fibres,

CC atrophy of upper cerebellar vermis; absence of Purkinje cells and
CC abnormal neuronal lipid storage. (I) can also be used to produce
CC antisense nucleic acids, is useful as molecular weight or chromosome
CC markers, to identify genetic disorders, as hybridisation probes or
CC primers, as an antigen, identify and express recombinant protein for
CC analysis, characterisation or therapeutic use, or as markers for tissues
CC in which the corresponding protein is expressed. Diagnostic methods from
CC the present invention can be used to identify subjects having or at risk
CC of developing a disease or disorder associated with aberrant expression
CC or activity of (I). The assays can be utilised to identify a subject
CC having or at risk of developing a disorder associated with Spastin
CC protein or spastin gene expression or activity. The present sequence
CC encodes a mutated human spastin.
CC N.B. The present sequence is not given in the present specification but
CC is derived from the human spastin nucleotide sequence (AAH20174) as
CC stated on page 14.
CC
XX
SQ Sequence 12793 BP; 4163 A; 2257 C; 2487 G; 3886 T; 0 other;

Query Match 100.0%; Score 20; DB 22; Length 12793;
Best Local Similarity 100.0%; Pred. No. 0.95;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtgaatggcaccattgcact 20
|||||
Db 6473 gtgaatggcaccattgcact 6492

RESULT 7
AAH20175
ID AAH20175 standard; DNA: 11493 BP.
XX
AC AAH20175;
XX
DT 09-AUG-2001 (first entry)
XX
DE Mouse spastin nucleotide sequence SEQ ID NO:3.
XX
KW Human; mouse; spastin; ARSACS; chromosome 13q11; identification;
KW autosomal recessive spastic ataxia of Charlevoix-Saguenay; mutation;
KW neurodegenerative disease; reduced sensory nerve conduction; diagnosis;
KW reduced motor nerve velocity; hypermyelination of retinal nerve fibre;
KW atrophy of upper cerebellar vermis; absence of Purkinje cell;
KW abnormal neuronal lipid storage; genetic disorder; characterisation; ds.
XX
OS Mus musculus.
XX
FH Key Location/Qualifiers
FT CDS 1..11493
FT /*tag= a
FT /product= "spastin"
XX
PN MO200129266-A2.
XX
PD 26-APR-2001.
XX
PF 20-OCT-2000; 2000WO-US29130.
XX
PR 20-OCT-1999; 99US-0160588.
XX
PA (UYMC-) UNIV MCGILL.
PA (HOPIT-) HOPITAL SAINTE-JUSTINE.
XX
PI Hudson TJ, Engert J, Richter A;
XX
DR WPI: 2001-308494/32.
DR P-PSDB; AAB97820.
XX
PT New isolated polynucleotide, encoding spastin gene, and polypeptides,
PT useful for diagnosing autosomal recessive spastic ataxia of
PT Charlevoix-Saguenay disease by detecting two point mutations in spastin
PT gene sequence -

XX
PS Claim 1; Fig 8; 76pp; English.
XX
CC The present invention describes human and mouse spastin, and mutated
CC human spastin (autosomal recessive spastic ataxia of Charlevoix-Saguenay
CC (ARSACS)) gene sequences (I). The spastin gene has been mapped to
CC chromosome 13q11. (I) have neuroprotective activities and can be used in
CC gene therapy and as a spastin polypeptide agonists. (I), their fragments
CC or their complements can be useful for assaying the presence of a nucleic
CC acid molecule in a sample. (I) is useful for diagnosing or aiding in the
CC diagnosis of an early onset neurodegenerative disease in an individual.
CC The neurodegenerative disease comprises reduced sensory nerve conduction,
CC reduced motor nerve velocity, hypermyelination of retinal nerve fibres,
CC atrophy of upper cerebellar vermis, absence of Purkinje cells and
CC abnormal neuronal lipid storage. (I) can also be used to produce
CC antisense nucleic acids, is useful as molecular weight or chromosome
CC markers, to identify genetic disorders, as hybridisation probes or
CC primers, as an antigen, identify and express recombinant protein for
CC analysis, characterisation or therapeutic use, or as markers for tissues
CC in which the corresponding protein is expressed. Diagnostic methods from
CC the present invention can be used to identify subjects having or at risk
CC of developing a disease or disorder associated with aberrant expression
CC or activity of (I). The assays can be utilised to identify a subject
CC having or at risk of developing a disorder associated with Spastin
CC protein or spastin gene expression or activity. The present sequence
CC encodes mouse spastin as given in the present invention.
CC
XX
SQ Sequence 11493 BP; 3599 A; 2281 C; 2387 G; 3226 T; 0 other;

Query Match 92.0%; Score 18.4; DB 22; Length 11493;
Best Local Similarity 95.0%; Pred. No. 6.7;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gtgaatggcaccattgcact 20
|||||
Db 6397 gtgaatggcaccattgcact 6416

RESULT 8
ABA08683
ID ABA08683 standard; cDNA: 901 BP.
XX
AC ABA08683;
XX
DT 11-JAN-2002 (first entry)
XX
DE Human PRO257 homologue-encoding cDNA, SEQ ID NO:459.
XX
KW Human; cytokine; cell proliferation; cell differentiation; growth factor;
KW haematopoiesis regulation; tissue growth; immunomodulator; actvlin;
KW inhibin; chemotaxis; chemokinesis; thrombolytic; oncogenesis;
KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;
KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;
KW chronic inflammatory condition; proliferative retinopathy;
KW atherosclerosis; coronary heart disease; arterial ischaemia;
KW bone disorder; osteoporosis; vascular growth disorder;
KW tissue regeneration; wound healing; infection; immune disorder;
KW cell culture; drug screening; gene therapy; antiinflammatory;
KW antiasthmatic; antiarthritis; haemostatic; antiarteriosclerotic;
KW cytoskeletal; osteopathic; vasotropic; cardiant; virucide; antibacterial;
KW antifungal; vulnery; antitumor; ss.
XX
OS Homo sapiens.
XX
PN MO200157188-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US03800.
XX
PR 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KM food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI: 2001-639362/73.
DR P-PSDB: ABG15064.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID No 15055; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.
XX
XX Sequence 2206 BP; 573 A; 638 C; 481 G; 514 T; 0 other;

Query Match 82.0%; Score 16.4; DB 23; Length 2206;
Best Local Similarity 94.4%; Pred. No. 58;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gtgaatgcccacttgcga 18
|||||
Db 2045 GTGAATGCCACTCTGCA 2028

RESULT 11
AAS82805/c
ID AAS82805 standard; CDNA: 2206 BP.
XX
AC AAS82805;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #18609.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI: 2001-639362/73.
DR P-PSDB: ABG18618.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID No 18609; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.
XX
XX Sequence 2206 BP; 573 A; 638 C; 481 G; 514 T; 0 other;

Query Match 82.0%; Score 16.4; DB 23; Length 2206;
Best Local Similarity 94.4%; Pred. No. 58;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gtgaatgcccacttgcga 18
|||||
Db 2045 GTGAATGCCACTCTGCA 2028

RESULT 12
AAS83970/c
ID AAS83970 standard; CDNA: 2206 BP.
XX
AC AAS83970;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #19774.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX Homo sapiens.
 OS
 XX
 XX WO200175067-A2.
 PN
 XX
 PD 11-OCT-2001.
 XX
 XX 30-MAR-2001; 2001WO-US08631.
 PF
 XX 31-MAR-2000; 2000US-0540217.
 PR
 XX 23-AUG-2000; 2000US-0649167.
 PA
 XX (HYSE-) HYSEQ INC.
 PI
 XX Drmanac RT, Liu C, Tang YT;
 DR
 XX WPI; 2001-639362/73.
 DR
 XX P-PSDB; ABG19783.
 XX
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 PS
 PS Claim 1; SEQ ID No 19774; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 XX Sequence 2206 BP; 573 A; 638 C; 481 G; 514 T; 0 other;

Query Match 82.0%; Score 16.4; DB 23; Length 2206;
 Best Local Similarity 94.4%; Pred. No. 58;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 gtaatgacacttgcga 18
 |||||
 Db 2045 GTGAATGCCACTCTGCA 2028

RESULT 13
 AAS85113/c
 ID AAS85113 standard; cDNA; 2206 BP.
 XX
 AC AAS85113;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE DNA encoding novel human diagnostic protein #20917.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX

OS Homo sapiens.
 XX
 XX WO200175067-A2.
 PN
 XX
 PD 11-OCT-2001.
 XX
 XX 30-MAR-2001; 2001WO-US08631.
 PF
 XX 31-MAR-2000; 2000US-0540217.
 PR
 XX 23-AUG-2000; 2000US-0649167.
 PA
 XX (HYSE-) HYSEQ INC.
 PI
 XX Drmanac RT, Liu C, Tang YT;
 DR
 XX WPI; 2001-639362/73.
 DR
 XX P-PSDB; ABG20926.
 XX
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 PS
 PS Claim 1; SEQ ID No 20917; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 XX Sequence 2206 BP; 573 A; 638 C; 481 G; 514 T; 0 other;

Query Match 82.0%; Score 16.4; DB 23; Length 2206;
 Best Local Similarity 94.4%; Pred. No. 58;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 gtaatgacacttgcga 18
 |||||
 Db 2045 GTGAATGCCACTCTGCA 2028

RESULT 14
 AAS88179/c
 ID AAS88179 standard; cDNA; 2206 BP.
 XX
 AC AAS88179;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE DNA encoding novel human diagnostic protein #23983.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX
 OS Homo sapiens.

XX MO200175067-A2.
PN 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US08631.
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX (HYSE-) HYSEQ INC.
PA Drmanac RT, Liu C, Tang YT;
PI WPI: 2001-639362/73.
XX P-PSDB; ABG23992.
DR
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX Claim 1; SEQ ID No 23983; 103pp; English.
PS
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2206 BP; 573 A; 638 C; 481 G; 514 T; 0 other;

Query Match 82.0%; Score 16.4; DB 23; Length 2206;
Best Local Similarity 94.4%; Pred. No. 58;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gtgaatggccacttgc 18
|||
Db 2045 GTGAATGGCCACTCTGCA 2028

RESULT 15
ID AAS89592/c
XX AAS89592 standard; cDNA; 2206 BP.
AC AAS89592;
XX
XX 13-FEB-2002 (first entry)
XX
XX DNA encoding novel human diagnostic protein #25396.
DE
XX
XX Human: chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX Homo sapiens.
XX

PN MO200175067-A2.
XX 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US08631.
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX (HYSE-) HYSEQ INC.
PA Drmanac RT, Liu C, Tang YT;
PI WPI: 2001-639362/73.
XX P-PSDB; ABG25405.
DR
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX Claim 1; SEQ ID No 25396; 103pp; English.
PS
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2206 BP; 573 A; 638 C; 481 G; 514 T; 0 other;

Query Match 82.0%; Score 16.4; DB 23; Length 2206;
Best Local Similarity 94.4%; Pred. No. 58;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gtgaatggccacttgc 18
|||
Db 2045 GTGAATGGCCACTCTGCA 2028

Search completed: May 22, 2002, 06:46:09
Job time: 8179 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 22, 2002, 06:34:29 ; Search time 108.82 Seconds
(without alignments)
45.145 Million cell updates/sec

Title: US-09-693-205-47

Perfect score: 20
Sequence: 1 gtgaatgccactgtcact 20

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents.NA:*
1: /cgn2_6/pdata/1/lna/5A_COMB.seq:*
2: /cgn2_6/pdata/1/lna/5B_COMB.seq:*
3: /cgn2_6/pdata/1/lna/6A_COMB.seq:*
4: /cgn2_6/pdata/1/lna/6B_COMB.seq:*
5: /cgn2_6/pdata/1/lna/6C_COMB.seq:*
6: /cgn2_6/pdata/1/lna/6D_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	15.8	79.0	1445	1	US-08-324-533-1	Sequence 1, Appl
2	15.4	77.0	5105	1	US-08-148-122A-1	Sequence 1, Appl
3	15.2	76.0	768	2	US-08-799-464A-2	Sequence 2, Appl
4	15.2	76.0	768	5	PCR-US95-09927-2	Sequence 2, Appl
5	15.2	76.0	1304	6	5179023-3	Patent No. 5179023
6	15.2	76.0	2351	4	US-08-478-316-4	Sequence 4, Appl
7	15.2	76.0	2352	4	US-08-478-316-5	Sequence 5, Appl
8	15.2	76.0	2352	4	US-08-478-316-7	Sequence 7, Appl
9	15.2	76.0	3293	4	US-08-478-316-40	Sequence 40, Appl
10	15.2	76.0	3293	4	US-08-478-316-41	Sequence 41, Appl
11	15.2	76.0	3285	4	US-08-686-968C-5	Sequence 5, Appl
12	15.2	76.0	3358	2	US-08-799-464A-1	Sequence 1, Appl
13	15.2	76.0	3358	2	PCR-US95-09927-1	Sequence 1, Appl
14	14.8	74.0	3342	2	US-08-254-989-1	Sequence 1, Appl
15	14.8	74.0	5596	4	US-09-078-294-5	Sequence 5, Appl
16	14.8	74.0	80246	4	US-09-078-294-4	Sequence 4, Appl
17	14.8	74.0	80595	4	US-09-078-294-3	Sequence 3, Appl
18	14.4	72.0	2574	4	US-09-142-529-2	Sequence 2, Appl
19	14.2	71.0	551	2	US-08-647-368A-4	Sequence 4, Appl
20	14.2	71.0	1194	4	US-08-943-731-188	Sequence 188, App
21	14.2	71.0	2205	4	US-08-840-767-53	Sequence 53, Appl
22	14.2	71.0	2449	4	US-08-840-767-7	Sequence 7, Appl
23	14.2	71.0	3660	1	US-08-158-232-42	Sequence 42, Appl
24	14.2	71.0	3660	1	US-08-611-928-42	Sequence 42, Appl
25	14.2	71.0	3660	3	US-09-173-891-42	Sequence 42, Appl
26	14.2	71.0	3796	1	US-08-343-760A-1	Sequence 1, Appl
27	14.2	71.0	20084	4	US-08-943-731-5	Sequence 5, Appl

28	14	70.0	1716	5	PCR-US96-05320A-541	Sequence 541, App
29	13.8	69.0	1236	3	US-09-081-320-37	Sequence 37, Appl
30	13.8	69.0	1245	1	US-08-385-186-14	Sequence 14, Appl
31	13.8	69.0	1247	2	US-08-647-960-1	Sequence 1, Appl
32	13.8	69.0	1405	2	US-08-555-723B-5	Sequence 5, Appl
33	13.8	69.0	1405	3	US-09-123-465-5	Sequence 1, Appl
34	13.8	69.0	1419	3	US-08-726-807B-1	Sequence 1, Appl
35	13.8	69.0	1419	3	US-09-258-367-1	Sequence 1, Appl
36	13.8	69.0	1419	4	US-09-546-550-1	Sequence 1, Appl
37	13.8	69.0	1419	4	US-09-431-414-1	Sequence 1, Appl
38	13.8	69.0	1419	4	US-09-225-670-1	Sequence 1, Appl
39	13.8	69.0	1425	1	US-08-385-186-12	Sequence 12, Appl
40	13.8	69.0	1710	4	US-08-630-915A-189	Sequence 189, App
41	13.8	69.0	1978	1	US-08-614-801A-3	Sequence 3, Appl
42	13.8	69.0	2211	4	US-08-891-298-2	Sequence 2, Appl
43	13.8	69.0	2797	2	US-08-555-723B-4	Sequence 4, Appl
44	13.8	69.0	2797	3	US-09-123-465-4	Sequence 4, Appl
45	13.8	69.0	3892	2	US-08-555-723B-3	Sequence 3, Appl

ALIGNMENTS

RESULT 1
US-08-324-533-1/c
; Sequence 1, Application US/08324533
; Patent No. 5547870
; GENERAL INFORMATION:
; APPLICANT: Delta, Asis
; APPLICANT: Mehta, Anuradha
; APPLICANT: Nataraajan, K.
; TITLE OR INVENTION: Oxalate Decarboxylase
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cushman, Darby & Cushman
; STREET: 1100 New York Avenue, N.W.
; City: Washington
; STATE: D.C.
; COUNTRY: U.S.
; ZIP: 20005-3918
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/324,533
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/985,695
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Kokulis, Paul N.
; REGISTRATION NUMBER: 16773
; TELEPHONE: 202-861-3000
; TELEFAX: 202-822-0944
; TELEX: 6714627 cush
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1445 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-324-533-1

Query Match 79.0%; Score 15.8; DB 1; Length 1445;
Best local Similarity 89.5%; Pred. No. 13;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 tgaatggcacttgcact 20
|||||
DB 345 TGAATGGCCTTGCCTTGCCTT 327

RESULT 2
US-08-148-122A-1/C

; Sequence 1, Application US/08148122A
; Patent No. 5441885

GENERAL INFORMATION:

APPLICANT: Goldberg, Ina
APPLICANT: Allenza, Paul
TITLE OF INVENTION: IMPROVED BACTERIAL STRAINS FOR
TITLE OF INVENTION: BIOREMEDIATION
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: AlliedSignal Inc.
STREET: 50 East Algonquin Road
CITY: Des Plaines
STATE: Illinois
COUNTRY: U.S.A.
ZIP: 60017-5016

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5 inch diskette
COMPUTER: AST Bravo 486/33 DX
OPERATING SYSTEM: DOS 5.00
SOFTWARE: Microsoft Word 2.0c

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/148,122A
FILING DATE: No. 544185ember 5, 1993
CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/049,281
FILING DATE: April 13, 1993

ATTORNEY/AGENT INFORMATION:

NAME: Mary Jo Boldingh
REGISTRATION NUMBER: 34,713
TELECOMMUNICATION INFORMATION:
TELEPHONE: (708) 391-3383
TELEFAX: (708) 391-3356

TELEX: N/A

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 5105

TYPE: nucleic acid

STRANDEDNESS: double strand

TOPOLOGY: linear

US-08-148-122A-1

Query Match 77.0%; Score 15.4; DB 1; Length 5105;
Best Local Similarity 94.1%; Pred. No. 27;

Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 aatggcacttgcact 20
|
DB 4880 AGTGGCCTTGCCTTGCCTTGCCTT 4864

RESULT 3
US-08-799-464A-2/C

; Sequence 2, Application US/08799464A
; Patent No. 598601

GENERAL INFORMATION:

APPLICANT: Murtough, Michael P. et al.
TITLE OF INVENTION: VR-2332 VIRAL NUCLEOTIDE SEQUENCE AND
TITLE OF INVENTION: METHODS OF USE
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: John M. Collins
STREET: 2405 Grand Blvd., Suite 400

CITY: Kansas City
STATE: Missouri
COUNTRY: USA
ZIP: 64108

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/799,464A

FILING DATE:

CLASSIFICATION: 536

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/287,941

FILING DATE: August 5, 1994

ATTORNEY/AGENT INFORMATION:

NAME: Collins, John M.
REGISTRATION NUMBER: 26,262
REFERENCE/DOCKET NUMBER: 22907
TELECOMMUNICATION INFORMATION:
TELEPHONE: (816) 474-9050
TELEFAX: (816) 474-9057

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 768 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: CDNA

FEATURE:

NAME/KEY: CDS

LOCATION: 1..768

IDENTIFICATION METHOD: experimental

OTHER INFORMATION: /evidence= EXPERIMENTAL

US-08-799-464A-2 /standard_name= "VR-2332 ORF 2"

Query Match 76.0%; Score 15.2; DB 2; Length 768;
Best Local Similarity 85.0%; Pred. No. 23;

Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 gtaatggcacttgcact 20
|
DB 261 GGAATGCTCCTTGCCTTGCCTTGCCTT 242

RESULT 4
PCT-US95-09927-2/C

; Sequence 2, Application PC/TUS9509927
; GENERAL INFORMATION:

APPLICANT: Murtough, Michael P.
TITLE OF INVENTION: VR-2332 VIRAL NUCLEOTIDE SEQUENCE AND

TITLE OF INVENTION: METHODS OF USE
NUMBER OF SEQUENCES: 26

CORRESPONDENCE ADDRESS:
ADDRESSEE: John M. Collins

STREET: 1101 Walnut, Suite 1400

CITY: Kansas City

STATE: Missouri

COUNTRY: USA

ZIP: 64106

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/09927

FILING DATE:

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

```

1 NAME: Collins, John M.
2 REGISTRATION NUMBER: 26122
3 REFERENCE/DOCKET NUMBER: 22907
4 TELECOMMUNICATION INFORMATION:
5 TELEPHONE: (816) 474-9050
6 TELEFAX: (816) 474-9057
7 INFORMATION FOR SEQ. ID NO. 2:
8 SEQUENCE CHARACTERISTICS:
9 LENGTH: 768 base pairs
10 TYPE: nucleic acid
11 STRANDEDNESS: double
12 TOPOLOGY: linear
13 MOLECULE TYPE: cDNA
14 FEATURE:
15 NAME/KEY: CDS
16 LOCATION: 1..768
17 IDENTIFICATION METHOD: experimental
18 OTHER INFORMATION: /evidence= EXPERIMENTAL
19 OTHER INFORMATION: /standard_name="VR-2332 ORF 2"
20 /cct=US95-09921-2

```

Query Match	76.0%;	Score 15.2;	DB 5;	length 768;
Best Local Similarity	85.0%;	Pred. No. 23;		
Matches 17; Conservative	0;	Mismatches 3;	Indels 0;	Gaps 0

```

QY      1 gtgaatgcccactttgcac 20
          | | | | | | | | | |
Db      261 GGGATGTCACACTTGGCACT 242

```

RESULT 5
5179023-3
Patent No. 5179023
APPLICANT: CALHOUN, DAVID H.; COPPOLA, GEORGE
TITLE OF INVENTION: RECOMBINANT a-GALACTOSIDASE A THERAPY
FOR FABRY DISEASE
NUMBER OF SEQUENCES: 4
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/520,312
FILING DATE: 07-MAY-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 328,421
FILING DATE: 24-MAR-1989
SEQ ID NO.: 3
LENGTH: 1304
5179023-3

Query Match	76.0%;	Score 15.2;	DB 6;	Length 1304;
Best Local Similarity	85.0%;	Pred. No. 26;		
Matches 17; Conservative	0;	Mismatches 3;	Indels 0;	Gaps 0

```

QY      1  gtgaatgccaacttgcact 20
          |||||  |||||  ||
Db      807  gtgaatgccaacttggcct 826

```

RESULT 6
 US-08-478-316-4/c
 ; Sequence 4, Application US/08478316
 ; Patent No. 6251397
 ;
 ; GENERAL INFORMATION:
 ;
 ; APPLICANT: PAUL, PREM S.
 ; APPLICANT: HALBUR, PATRICK G.
 ;
 ; APPLICANT: MENG, XIANG-JIN
 ;
 ; APPLICANT: MORZOV, IGOR
 ;
 ; TITLE OF INVENTION: POLYNUCLEIC ACIDS ISOLATED FROM A PORCINE
 ;
 ; TITLE OF INVENTION: REPRODUCTIVE AND RESPIRATORY SYNDROME VIRUS (PRRSV), PROTEINS
 ;
 ; TITLE OF INVENTION: ENCODED BY THE POLYNUCLEIC ACIDS, VACCINES BASED ON THE
 ;
 ; TITLE OF INVENTION: PROTEINS AND/OR POLYNUCLEIC ACIDS, A METHOD OF PROTECTING A
 ;
 ; TITLE OF INVENTION: PIG FROM PRRS AND A METHOD OF DETECTING A PRRSV
 ;
 ; NUMBER OF SEQUENCES: 95

1 CORRESPONDENCE ADDRESS:
2 ADDRESSEE: OBION, SPIVAK, McCLELLAND, MAIER & NEUSTADT
3 ADDRESSEE: P. C.
4 STREET: 1755 S. Jefferson Davis Highway, Suite 400
5 CITY: Arlington
6 STATE: Virginia
7 COUNTRY: U.S.A.

```

? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? OPERATING SYSTEM: IBM PC compatible
? SOFTWARE: PC-DOS/MS-DOS
? CURRENT APPLICATION DATA: Patentin Release #1.0, Version #1.30
? APPLICATION NUMBER: US/08/478,316
? FILING DATE: 07-JUN-1995

```

APPLICATION NUMBER: US 08/301,435
FILING DATE: 01-SEP-1994

INVENTION NUMBER: 05-00-1221-02
FILING DATE: 05-OCT-1993
PRIOR APPLICATION DATA:

; FILING DATE: 30-OCT-1992
 ;
 ; ATTORNEY/AGENT INFORMATION:
 ;

REFERENCE/DOCKET NUMBER: 4625-026-55X CIP
TELECOMMUNICATION INFORMATION:

TELEX: 248855 OPAT UR
INFORMATION FOR SEO ID NO: 4:

LENGTH: 2501 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown

```

; MOLECULE TYPE:  CDNA
; ORIGINAL SOURCE:
;

```

INDIVIDUAL ISOLATE: ISU222
US-08-478-316-4

Query Match	76.0%	Score 15.2	DB 4	Length 2351
Best Local Similarity	85.0%	Pred. No. 30		
Matches 17; Conservative	0	Mismatches 3	Indels 0	Gaps 0

```

QY      1  gtgaatgccaatttgcaat  20
          | | | | | | | | | |
Db      295 GGGATGTCACCTTGGCACT  276

```

RESULT 7
 US-08-478-316-5/C
 : Sequence 5, Application US/08478316
 : Patent No. 6251397
 :
 : GENERAL INFORMATION:
 :
 : APPLICANT: PAUL, PREM S.
 :
 : APPLICANT: HALBOR, PATRICK G.
 :
 : APPLICANT: MENG, XIANG-JIN
 :
 : APPLICANT: MOROZOV, ISOR
 :
 : TITLE OF INVENTION: POLYNUCLEIC ACIDS ISOLATED FROM A PORCINE
 :
 : TITLE OF INVENTION: REPRODUCTIVE AND RESPIRATORY SYNDROME VIRUS (PRRSV), PROTEIN
 :
 : TITLE OF INVENTION: ENCODED BY THE POLYNUCLEIC ACIDS, VACCINES BASED ON THE
 :
 : TITLE OF INVENTION: PROTEINS AND/OR POLYNUCLEIC ACIDS, A METHOD OF PROTECTING A
 :
 : TITLE OF INVENTION: PIG FROM PRRS AND A METHOD OF DETECTING A PRRSV
 :
 : NUMBER OF SEQUENCES: 95
 :
 : CORRESPONDENCE ADDRESSES:
 :

ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
ADDRESSEE: P.C.
STREET: 1755 S. Jefferson Davis Highway, Suite 400
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/478,316
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/301,435
FILING DATE: 01-SEP-1994
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/131,625
FILING DATE: 05-OCT-1993
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/969,071
FILING DATE: 30-OCT-1992
ATTORNEY/AGENT INFORMATION:
NAME: Lavalleye, Jean-Paul M.P.
REGISTRATION NUMBER: 31,451
REFERENCE/DOCKET NUMBER: 4625-026-55X CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 413-3000
TELEFAX: (703) 413-2220
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2352 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: CDNA
ORIGINAL SOURCE:
ORGANISM: Porcine reproductive and respiratory virus
INDIVIDUAL ISOLATE: VR2332
US-08-478-316-5

Query Match 76.0%; Score 15.2; DB 4; Length 2352;
Best Local Similarity 85.0%; Pred. No. 30;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 gtgaatgcccacttgcact 20
| ||||| ||||| |||||
Db 295 GGAATGTCCACTTGGCAGT 276

RESULT 8
US-08-478-316-7/C
Sequence 7, Application US/08478316
Patent No. 6251397
GENERAL INFORMATION:
APPLICANT: PAUL, PREM S.
APPLICANT: HALBUR, PATRICK G.
APPLICANT: MENG, XIANG-JIN
APPLICANT: MOROZOV, IGOR
TITLE OF INVENTION: POLYNUCLEIC ACIDS ISOLATED FROM A PORCINE
REPRODUCTIVE AND RESPIRATORY SYNDROME VIRUS (PRRSV), PROTEINS
TITLE OF INVENTION: ENCODED BY THE POLYNUCLEIC ACIDS, VACCINES BASED ON THE
TITLE OF INVENTION: PROTEINS AND/OR POLYNUCLEIC ACIDS, A METHOD OF PROTECTING A
TITLE OF INVENTION: PIG FROM PRRS AND A METHOD OF DETECTING A PRRSV
NUMBER OF SEQUENCES: 95
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
ADDRESSEE: P.C.
STREET: 1755 S. Jefferson Davis Highway, Suite 400
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202

STREET: 1755 S. Jefferson Davis Highway, Suite 400
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/478,316
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/301,435
FILING DATE: 01-SEP-1994
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/131,625
FILING DATE: 05-OCT-1993
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/969,071
FILING DATE: 30-OCT-1992
ATTORNEY/AGENT INFORMATION:
NAME: Lavalleye, Jean-Paul M.P.
REGISTRATION NUMBER: 31,451
REFERENCE/DOCKET NUMBER: 4625-026-55X CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 413-3000
TELEFAX: (703) 413-2220
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2352 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: CDNA
ORIGINAL SOURCE:
ORGANISM: Porcine reproductive and respiratory virus
INDIVIDUAL ISOLATE: ISU79
US-08-478-316-7

Query Match 76.0%; Score 15.2; DB 4; Length 2352;
Best Local Similarity 85.0%; Pred. No. 30;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 gtgaatgcccacttgcact 20
| ||||| ||||| |||||
Db 295 GGAATGTCCACTTGGCAGT 276

RESULT 9
US-08-478-316-40/C
Sequence 40, Application US/08478316
Patent No. 6251397
GENERAL INFORMATION:
APPLICANT: PAUL, PREM S.
APPLICANT: HALBUR, PATRICK G.
APPLICANT: MENG, XIANG-JIN
APPLICANT: MOROZOV, IGOR
TITLE OF INVENTION: POLYNUCLEIC ACIDS ISOLATED FROM A PORCINE
REPRODUCTIVE AND RESPIRATORY SYNDROME VIRUS (PRRSV), PROTEI
TITLE OF INVENTION: ENCODED BY THE POLYNUCLEIC ACIDS, VACCINES BASED ON THE
TITLE OF INVENTION: PROTEINS AND/OR POLYNUCLEIC ACIDS, A METHOD OF PROTECTING A
TITLE OF INVENTION: PIG FROM PRRS AND A METHOD OF DETECTING A PRRSV
NUMBER OF SEQUENCES: 95
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
ADDRESSEE: P.C.
STREET: 1755 S. Jefferson Davis Highway, Suite 400
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202

Best Local Similarity 85.0%; Pred. No. 32;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 ggaatgccacttgacct 20
1 ||||| ||||| |||||
Db 307 GGAATGCCACTTGCGACT 288

RESULT 12
US-08-799-464A-1/c
Sequence 1, Application US/08799464A
Patent No. 5998601
GENERAL INFORMATION:
APPLICANT: Murtough, Michael P. et al.
TITLE OF INVENTION: VR-2332 VIRAL NUCLEOTIDE SEQUENCE AND
TITLE OF INVENTION: METHODS OF USE
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESSES:
ADDRESSEE: John M. Collins
STREET: 2405 Grand Blvd., Suite 400
CITY: Kansas City
STATE: Missouri
COUNTRY: USA
ZIP: 64108
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/799,464A
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/287,941
FILING DATE: August 5, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Collins, John M.
REGISTRATION NUMBER: 26,262
REFERENCE/DOCKET NUMBER: 22907
TELECOMMUNICATION INFORMATION:
TELEPHONE: (816) 474-9050
TELEFAX: (816) 474-9057
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3358 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Arteriviridae (Unclassified)
STRAIN: VR-2332
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1..768
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /standard_name= "VR-2332 ORF2"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 624..1385
OTHER INFORMATION: /standard_name= "VR-2332 ORF 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1169..1701
OTHER INFORMATION: /standard_name= "VR-2332 ORF 4"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1716..2315
OTHER INFORMATION: /standard_name= "VR-2332 ORF 5"

FEATURE:
NAME/KEY: misc_feature
LOCATION: 2303..2824
OTHER INFORMATION: /standard_name= "VR-2332 ORF 6"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 2817..3185
OTHER INFORMATION: /standard_name= "VR-2332 ORF 7"
US-08-799-464A-1

Query Match 76.0%; Score 15.2; DB 2; Length 3358;
Best Local Similarity 85.0%; Pred. No. 32;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 ggaatgccacttgacct 20
1 ||||| ||||| |||||
Db 261 GGAATGCCACTTGCGACT 242

RESULT 13
PCT-US95-09927-1/c
Sequence 1, Application PC/TUS9509927
GENERAL INFORMATION:
APPLICANT: Murtough, Michael P.
TITLE OF INVENTION: VR-2332 VIRAL NUCLEOTIDE SEQUENCE AND
TITLE OF INVENTION: METHODS OF USE
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESSES:
ADDRESSEE: John M. Collins
STREET: 1101 Walnut, Suite 1400
CITY: Kansas City
STATE: Missouri
COUNTRY: USA
ZIP: 64106
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/09927
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Collins, John M.
REGISTRATION NUMBER: 26122
REFERENCE/DOCKET NUMBER: 22907
TELECOMMUNICATION INFORMATION:
TELEPHONE: (816) 474-9050
TELEFAX: (816) 474-9057
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3358 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Arteriviridae (Unclassified)
STRAIN: VR-2332
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1..768
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /standard_name= "VR-2332 ORF2"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 624..1385
OTHER INFORMATION: /standard_name= "VR-2332 ORF 3"

```
;
;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1169..1701
; OTHER INFORMATION: /standard_name= "VR-2332 ORF 4"
;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1716..2315
; OTHER INFORMATION: /standard_name= "VR-2332 ORF 5"
;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2303..2824
; OTHER INFORMATION: /standard_name= "VR-2332 ORF 6"
;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2817..3185
; OTHER INFORMATION: /standard_name= "VR-2332 ORF 7"
;
; PCT-US95-09927-1

Query Match          76.0%; Score 15.2; DB 5; Length 3358;
Best Local Similarity 85.0%; Pred. No. 32;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 gtaatggcacttgac 20
Db 261 GGGATGCTCCACTTGGCAGT 242

RESULT 14
US-08-254-989-1
; Sequence 1, Application US/08254989
; Patent No. 5859307
;
; GENERAL INFORMATION:
; APPLICANT: Mombaerts, Peter
; APPLICANT: Tonegawa, Susumu
; APPLICANT: Johnson, Randall S.
; APPLICANT: Papalioannou, Virginia
; TITLE OF INVENTION: Mutant RAG-1 Deficient Animals Having NO
; TITLE OF INVENTION: Mature B and T Lymphocytes
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kilpatrick & Cody
; STREET: 1100 Peachtree Street, Suite 2800
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: U.S.
; ZIP: 30309-4530
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/254,989
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/830831
; FILING DATE: 04-FEB-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: M15783
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 815-6508
; TELEFAX: (404) 815-6555
; INFORMATION FOR SEQ. ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3342 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
```

```
;
;
; HYPOTHETICAL: NO
;
; ANTI-SENSE: NO
;
; ORIGINAL SOURCE:
;
; ORGANISM: Mouse
;
; US-08-254-989-1

Query Match          74.0%; Score 14.8; DB 2; Length 3342;
Best Local Similarity 88.9%; Pred. No. 53;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 tgaatggcacttgac 19
Db 2640 TGAATGGCAACTTGTGCC 2657

RESULT 15
US-09-078-294-5/C
; Sequence 5, Application US/09078294
; Patent No. 6265211
;
; GENERAL INFORMATION:
; APPLICANT: Choo, Kong-Hong Andy
; APPLICANT: Du Sart, Desiree
; APPLICANT: Cancilla, Michael R.
; TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
; FILE REFERENCE: Davies Col
; CURRENT APPLICATION NUMBER: US/09/078,294
; CURRENT FILING DATE: 1998-05-13
; NUMBER OF SEQ. ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ. ID NO 5
; LENGTH: 5596
; TYPE: DNA
; ORGANISM: BAC-F2 contig 1
;
; US-09-078-294-5

Query Match          74.0%; Score 14.8; DB 4; Length 5596;
Best Local Similarity 88.9%; Pred. No. 59;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtaatggcacttgac 18
Db 4457 GGGATGGCCTCTTGCA 4440
```

Search completed: May 22, 2002, 06:34:32
Job time: 7577 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 22, 2002, 05:31:36 ; Search time 3619.39 Seconds
(without alignments)
74.581 Million cell updates/sec

Title: US-09-693-205-47

Perfect score: 20

Sequence: 1 gtgaatggccacttgcact 20

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estlin:*
4: em_estlmu:*
5: em_estlov:*
6: em_estlpl:*
7: em_estlro:*
8: em_hlc:*
9: gb_estcl:*
10: gb_estl2:*
11: gb_hlc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20	100.0	238	9	AA897178 am09e08.s
2	20	100.0	1010	10	BM472954 AGENCOURT
3	20	100.0	1083	10	BM476887 AGENCOURT
4	18.4	92.0	621	12	AQ446973
5	18.4	92.0	705	12	AQ255211
6	17.4	87.0	299	10	BF944309
7	17.4	87.0	300	10	BF944296
8	17.4	87.0	306	10	BF952284
9	17.4	87.0	311	10	BF944321
10	17.4	87.0	313	10	BF944323
11	17.4	87.0	315	10	BF944314
12	17.4	87.0	316	10	BF952276
13	17.4	87.0	434	10	W60474
14	17.4	87.0	477	10	BB652591
15	17.4	87.0	610	12	A2295173
16	17.4	87.0	655	9	BB624324
17	16.8	84.0	199	9	BB072918

18	16.8	84.0	329	10	BG958614
19	16.8	84.0	601	10	BB624992
20	16.8	84.0	611	9	AA163498
21	16.8	84.0	618	9	AA410991
22	16.8	84.0	635	9	BB589330
23	16.8	84.0	639	10	BI557012
24	16.8	84.0	656	12	A2382874
25	16.8	84.0	712	9	AA916846
26	16.8	84.0	813	10	BG781638
27	16.8	84.0	825	10	BF261027
28	16.4	82.0	220	10	BG155981
29	16.4	82.0	581	10	BM334925
30	16.4	82.0	581	12	CT2388
31	16.4	82.0	593	10	BF263818
32	16.4	82.0	628	10	B1005815
33	16.4	82.0	654	12	BH384785
34	16.4	82.0	968	12	CNS00044A
35	16.4	82.0	425	12	AO583449
36	16	80.0	504	12	A2301231
37	16	80.0	627	12	AQ708956
38	16	80.0	730	12	BH537560
39	15.8	79.0	154	12	AQ907033
40	15.8	79.0	296	10	BG220942
41	15.8	79.0	297	9	AV086027
42	15.8	79.0	312	10	T86720
43	15.8	79.0	358	12	BH228144
44	15.8	79.0	364	10	BG509827
45	15.8	79.0	378	12	AQ923385

ALIGNMENTS

RESULT 1
AA897178 238 bp mRNA linear EST 04-JAN-1999
LOCUS am09e08.s1 Soares_NFL_T.GBC.S1 Homo sapiens cDNA clone
DEFINITION IMAGE:1466342 3', mRNA sequence.

ACCESSION AA897178
VERSION AA897178.1 GI:3033798
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
This clone is available royalty-free through LNL ; contact the IMAG Consortium (info@image.lnl.gov) for further information.
Insert Length: 847 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 132.
Location/Qualifiers
1. 238

FEATURES
source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1466342"
/clone_1lb="Soares_NFL_T.GBC.S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site.1: Not I; Site.2: Eco RI.
Equal amounts of plasmid DNA from three normalized libraries (fetal lung NBH19W, testis NHT, and B-cell NCI-CGAP_GCB1) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of

I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo. "

BASE COUNT 71 a 53 c 45 g 69 t

5', mRNA sequence.
ACCESSION BM476887
VERSION BM476887.1 GI:18525929
KEYWORDS EST
SOURCE human.
ORGANISM Homo sapiens

Query Match 100.0%; Score 20; DB 9; Length 238;
Best Local Similarity 100.0%; Pred. No. 9.1;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gtgaatggccactttgcact 20
|||||

REFERENCE NIH-MGC http://mgc.nci.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaabs-remail.nih.gov
Tissue Procurement: ATCC

RESULT 2
BM472954 1010 bp mRNA linear EST 05-FEB-2002
LOCUS AGENCOURT_6466106 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:5574588
DEFINITION 5', mRNA sequence.
ACCESSION BM472954
VERSION BM472954.1 GI:18521996
KEYWORDS EST
SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1010)

REFERENCE NIH-MGC http://mgc.nci.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaabs-remail.nih.gov
Tissue Procurement: ATCC

CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov

Plate: LLM12323 row: h column: 13
High quality sequence stop: 738.

FEATURES Location/Qualifiers

1..1010

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_image="5574588"

/clone_lib="NIH_MGC_88"

/tissue_type="quodanal adenocarcinoma, cell line"

/lab_host="DH10B (phage-resistant)"

/note="Organ: small intestine; Vector: pCMV-SPORT6;
Site_1: NotI; Site_2: SalI; Cloned unidirectionally;
Oligo-dT primed. Average insert size 1.767 kb. Library
enriched for full-length clones and constructed by Life
Technologies. Note: this is a NIH_MGC Library."

BASE COUNT 293 a 209 c 191 g 317 t

Query Match 100.0%; Score 20; DB 10; Length 1010;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gtgaatggccactttgcact 20
|||||

RESULT 4
A0446973 621 bp DNA linear GSS 08-APR-1999
LOCUS mgxb0002115f CUGI Rice Blast BAC Library Magnaporthe grisea genomic
DEFINITION clone mgxb0002115f, DNA sequence.
ACCESSION A0446973.1 GI:4576110
VERSION
KEYWORDS GSS.
SOURCE Magnaporthe grisea.
ORGANISM Magnaporthe grisea

BASE COUNT 293 a 209 c 191 g 317 t

REFERENCE NIH-MGC http://mgc.nci.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL
COMMENT Contact: Dean RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson University, Clemson, SC 29634
Tel: 864 656 5737
Fax: 864 656 4293
Email: rdean@clemson.edu
Seq primer: TAATAGACTACTATAGGG

Class: BAC ends
High quality sequence start: 52
High quality sequence stop: 286.

FEATURES

source

1. 621
Location/Qualifiers

/organism="Magnaporthe grisea"

/strain="70-15"

/db_xref="taxon:148305"

/clone="mgxb0002115f"

/clone_lib="CUGI Rice Blast BAC Library"

/tissue_type="Protoplasts"

/lab_host="E. coli DH10B"

/note="Vector: PBACWICH; Site 1: HindIII; Site 2: HindIII; Rice blast is one of the most devastating fungal diseases of rice world wide. It is a filamentous ascomycete with a haploid genome (n=7) of approximately 40 Mbp. Rice blast is an important model fungal pathogen for studying numerous aspects of the fungal-host interaction. In order to facilitate genome wide analysis, a BAC library containing 9216 clones with an average insert size of 130 kbp was constructed. This library represents greater than 25X genome coverage. High density colony filters are available upon request."

BASE COUNT

154 a 154 c 173 g 135 t 5 others

ORIGIN

Query Match

Best Local Similarity 92.0%; Score 18.4; DB 12; Length 621;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gtgaatgccacttgact 20

Db 414 GAGAAAGCCACTTGGCACT 433

RESULT 5

LOCUS

AQ255211 705 bp DNA linear GSS 23-OCT-1998
mgxb0009820r CUGI Rice Blast BAC Library Magnaporthe grisea genomic
clone mgxb0009820r. DNA sequence.

ACCESSION

AQ255211

AQ255211.1 GI:3779526

KEYWORDS

GSS.

Magnaporthe grisea.

SOURCE

Magnaporthe grisea.

ORGANISM

Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Sordariomycetes incertae sedis; Magnaporthaceae; Magnaporthe.

1 (bases 1 to 705)

Yu, Y., Zhu, H., Boyd, C.A., Gaudette, B., Gayle, A., Kingsbury, R.,
Phillips, K., Saslowski, M., Wing, R.A. and Dean, R.A.

A BAC End Sequencing Framework to Sequence the Magnaporthe grisea
Genome

Unpublished (1998)

JOURNAL

Contact: Dean RA

Clemson University Genomics Institute

100 Jordan Hall, Clemson University, Clemson, SC 29634

Tel: 864 656 5737

Fax: 864 656 4293

Email: rdean@clemson.edu

Seq primer: GGAACACGCTATGACCATG

Class: BAC ends

High quality sequence stop: 365.

FEATURES

source

1. 705

/organism="Magnaporthe grisea"

/strain="70-15"

/db_xref="taxon:148305"

/clone="mgxb0009820r"

/clone_lib="CUGI Rice Blast BAC Library"

/tissue_type="Protoplasts"

/lab_host="E. coli DH10B"

BASE COUNT

177 a 175 c 199 g 154 t

ORIGIN

Query Match 92.0%; Score 18.4; DB 12; Length 705;
Best Local Similarity 95.0%; Pred. No. 81;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gtgaatgccacttgact 20

Db 353 GAGAAAGCCACTTGGCACT 372

RESULT 6

LOCUS

BF944309 299 bp mRNA linear EST 22-JAN-2001
RC5-NN1163-131000-011-B11 NN1163 Homo sapiens cDNA, mRNA sequence.

ACCESSION

BF944309

BF944309.1 GI:12361584

KEYWORDS

EST.

SOURCE

human.

ORGANISM

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 299)

Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,

Nagal, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,

Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,

Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare

, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and

Simpson, A.J.

Shotgun sequencing of the human transcriptome with ORF expressed

sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

20202663

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, São Paulo-SP,

Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome

Project. This entry can be seen in the following URL

(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC56rc2-RC5-NN1163-

131000-011-B11&ts=2000-10-13&td=1)

Seq primer: puc 18 forward

High quality sequence stop: 299.

FEATURES

source

1. 299

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="NN1163"

/dev_stage="Adult"

/note="organ: nervous_normal; Vector: puc18; Site 1: SmaI;

Site 2: SmaI; A mini-library was made by cloning products

derived from ORFESTES PCR (U.S. Letters Patent application

No. 196,716 - Ludwig Institute for Cancer Research)

profiles into the puc 18 vector. Reverse transcription of

tissue mRNA and cDNA amplification were performed under

low stringency conditions."

BASE COUNT 68 a 87 c 93 g 51 t
ORIGIN

Query Match 87.0%; Score 17.4; DB 10; Length 299;
Best Local Similarity 94.7%; Pred. No. 1.9e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 tgaatggccactgtgcaact 20
|||||
Db 296 TGAATGGCCACTGTGCACT 278

RESULT 7
BF944296/c 300 bp mRNA linear EST 22-JAN-2001
LOCUS BF944296
DEFINITION RC5-NN1163-131000-011-B04 NN1163 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF944296
VERSION BF944296.1 GI:12361571
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS
1 (bases 1 to 300)
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M. R.,
Nagai, M. A., da Silva, W. Jr., Zago, M. A., Bordin, S., Costa, F. F.,
Goldman, G. H., Carvalho, A. F., Matsukuma, A., Bala, G. S., Simpson, D. H.,
Brunstein, A., deoliveira, P. S., Bucher, P., Jongeneel, C. V., O'Hare
M. J., Soares, F., Brentani, R. R., Reis, L. F., de Souza, S. J. and
Simpson, A. J.

TITLE
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags

JOURNAL
MEDLINE
COMMENT
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil

FEATURES
source
1.300
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1lb="NN1163"
/dev_stage="Adult"
/note="Organ: nervous normal; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
Location/Qualifiers
1.300

BASE COUNT 68 a 92 c 90 g 50 t
ORIGIN

Query Match 87.0%; Score 17.4; DB 10; Length 300;
Best Local Similarity 94.7%; Pred. No. 1.9e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 2 tgaatggccactgtgcaact 20
|||||

Db 291 TGAATGGCCACTGTGCACT 273

RESULT 8
BF952284/c 306 bp mRNA linear EST 22-JAN-2001
LOCUS BF952284
DEFINITION RC5-NN1163-101100-012-D07 NN1163 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF952284
VERSION BF952284.1 GI:12369559
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS
1 (bases 1 to 306)
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M. R.,
Nagai, M. A., da Silva, W. Jr., Zago, M. A., Bordin, S., Costa, F. F.,
Goldman, G. H., Carvalho, A. F., Matsukuma, A., Bala, G. S., Simpson, D. H.,
Brunstein, A., deoliveira, P. S., Bucher, P., Jongeneel, C. V., O'Hare
M. J., Soares, F., Brentani, R. R., Reis, L. F., de Souza, S. J. and
Simpson, A. J.

TITLE
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL
MEDLINE
COMMENT
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?rl=RC5&t2=RC5-NN1163-
101100-012-D07&t3=2000-11-10&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 7
High quality sequence stop: 305.
Location/Qualifiers
1.306

FEATURES
source
1.306
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1lb="NN1163"
/dev_stage="Adult"
/note="Organ: nervous normal; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
Location/Qualifiers
1.306

BASE COUNT 68 a 92 c 93 g 53 t
ORIGIN

Query Match 87.0%; Score 17.4; DB 10; Length 306;
Best Local Similarity 94.7%; Pred. No. 1.9e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 2 tgaatggccactgtgcaact 20
|||||
Db 287 TGAATGGCCACTGTGCACT 269

RESULT 9
BF944321/c 311 bp mRNA linear EST 22-JAN-2001
LOCUS BF944321
DEFINITION RC5-NN1163-131000-011-G11 NN1163 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF944321
VERSION BF944321.1 GI:12361596
KEYWORDS EST.

SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 311)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC5<2=RC5-NN1163-131000-011-G1<3=2000-10-13<4=1)
Seq primer: puc 18 forward
High quality sequence start: 17
High quality sequence stop: 311.
Location/Qualifiers
1. 311
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NN1163"
/dev_stage="Adult"
/note="Organ: nervous_normal; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
BASE COUNT 68 a 97 c 94 g 52 t
ORIGIN
Query Match 87.0%; Score 17.4; DB 10; Length 311;
Best Local Similarity 94.7%; Pred. No. 2e+02; 1; Indels 0; Gaps 0;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 2 tgaatggcacttgcact 20
|||||
Db 302 TGAATGCCACTGTGCCT 284
RESULT 10
BF944323 313 bp mRNA linear EST 22-JAN-2001
LOCUS RC5-NN1163-131000-011-H03 NN1163 Homo sapiens cDNA, mRNA sequence.
DEFINITION BF944323
ACCESSION BF944323.1 GI:12361598
VERSION
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 313)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and

Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC5<2=RC5-NN1163-131000-011-H03<3=2000-10-13<4=1)
Seq primer: puc 18 forward
High quality sequence start: 17
High quality sequence stop: 313.
Location/Qualifiers
1. 313
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NN1163"
/dev_stage="Adult"
/note="Organ: nervous_normal; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
BASE COUNT 69 a 93 c 96 g 55 t
ORIGIN
Query Match 87.0%; Score 17.4; DB 10; Length 313;
Best Local Similarity 94.7%; Pred. No. 2e+02; 1; Indels 0; Gaps 0;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 2 tgaatggcacttgcact 20
|||||
Db 304 TGAATGCCACTGTGCCT 286
RESULT 11
BF944314 315 bp mRNA linear EST 22-JAN-2001
LOCUS RC5-NN1163-131000-011-D08 NN1163 Homo sapiens cDNA, mRNA sequence.
DEFINITION BF944314
ACCESSION BF944314.1 GI:12361589
VERSION
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 315)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil

Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC5&t2=RC5-NN1163-131000-011-D08&t3=2000-10-13&t4=1)
 Seq primer: puc 18 forward
 High quality sequence start: 2
 High quality sequence stop: 315.
 Location/Qualifiers
 1. 315
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_1lb="NN1163"
 /dev_stage="Adult"
 /note="Organ: nervous, normal; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 54 a 96 c 92 g 73 t

ORIGIN

Query Match 87.0%; Score 17.4; DB 10; Length 315;
 Best Local Similarity 94.7%; Pred. No. 2e+02;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 tgaatggccacttgcact 20
 ||||| ||||| |||||
 Db 21 TGAATGCCACTGTGCAC 39

RESULT 12
 BF952276 316 bp mRNA linear EST 22-JAN-2001
 LOCUS RC5-NN1163-101100-012-D08 NN1163 Homo sapiens cDNA, mRNA sequence.
 DEFINITION BF952276
 ACCESSION BF952276.1 GI:12369551
 VERSION EST.
 KEYWORDS human.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 316)
 Dias Neto, E., Garcia Correa, R., Verjowski-Almeida, S., Briones, M. R., Nagai, M. A., da Silva, W. Jr., Zago, M. A., Bordin, S., Costa, F. F., Goldman, G. H., Carvalho, A. F., Matsukuma, A., Baia, G. S., Simpson, D. H., Brunstein, A., deoliveira, P. S., Bucher, P., Jongeneel, C. V., O'Hare, M. J., Soares, F., Brentani, R. R., Reis, L. F., de Souza, S. J. and Simpson, A. J.
 Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 20202663
 Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC5&t2=RC5-NN1163-101100-012-D08&t3=2000-11-10&t4=1)
 Seq primer: puc 18 forward
 High quality sequence start: 69
 High quality sequence stop: 316.

TITLE
 JOURNAL
 MEDLINE
 COMMENT

FEATURES
 source
 Location/Qualifiers
 1. 316
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_1lb="NN1163"
 /dev_stage="Adult"
 /note="Organ: nervous, normal; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 74 a 93 c 97 g 52 t

ORIGIN

Query Match 87.0%; Score 17.4; DB 10; Length 316;
 Best Local Similarity 94.7%; Pred. No. 2e+02;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 tgaatggccacttgcact 20
 ||||| ||||| |||||
 Db 307 TGAATGCCACTGTGCAC 289

RESULT 13
 W60474 434 bp mRNA linear EST 07-JUN-1996
 LOCUS zc98d08.s1 Pancreatic Islet Homo sapiens cDNA IMAGE:339183
 DEFINITION 3', mRNA sequence.
 ACCESSION W60474
 VERSION W60474.1 GI:1367235
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 434)
 Hillier, L., Lennon, G., Becker, M., Bonaldo, M. F., Chapell, B., Chisoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W., Hawkins, M., Hultman, M., Kucaba, T., Lacey, M., Le, M., Le, N., Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T., Schellenberg, K., Soares, M. B., Tan, F., Thierry-Mieg, J., Trevaskis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R. and Marra, M.
 Generation and analysis of 280,000 human expressed sequence tags
 Genome Res. 6 (9), 807-828 (1996)
 97044478
 Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available royalty-free through LNLN; contact the IMAGE Consortium (info@image.lnl.gov) for further information.
 Seq primer: mob.RCGA+ET
 High quality sequence stop: 340.
 Location/Qualifiers
 1. 434
 /organism="Homo sapiens"
 /db_xref="GDB:1264557"
 /db_xref="taxon:9606"
 /clone_1lb="IMAGE:339183"
 /clone_1lb="Pancreatic Islet"
 /tissue_type="Pancreatic Islet"
 /lab_host="SOLR cells (kanamycin resistant)"
 /note="Organ: pancreas; Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; Reference: Hum Mol Gen 2, 1795 (1993)
 Takeda et al. Cloned unidirectionally. Primer: 01190 dt.
 -5' adaptor sequence: 5' GAATTCGCGACGAG 3' -3' adaptor sequence: 5' CTCGAGCTTTTATTTTATTTT 3' "

BASE COUNT 98 a 93 c 97 g 146 t

TITLE
 JOURNAL
 MEDLINE
 COMMENT

ORIGIN

Query Match 87.0%; Score 17.4; DB 10; Length 434;
 Best Local Similarity 94.7%; Pred. No. 2.2e+02;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 2 tgaatgaccttgcact 20
 |||||
 Db 301 TGAATGCCACTTGCAC 319

RESULT 14
 BE652591/c 477 bp mRNA linear EST 06-SEP-2000
 LOCUS

DEFINITION
 UI-M-A00-ach-d-01-0-UI.r1 NIH_BMAP_MPG Mus musculus CDNA clone
 UI-M-A00-ach-d-01-0-UI 5', mRNA sequence.

ACCESSION
 BE652591
 VERSION
 BE652591.1 GI:9978440

KEYWORDS
 EST.

SOURCE
 house mouse.

ORGANISM
 Mus musculus

REFERENCE
 AUTHORS
 TITLE
 Normalization and subtraction: two approaches to facilitate gene discovery

JOURNAL
 MEDLINE
 97044477
 Contact: Chin, H

COMMENT
 National Institute of Mental Health
 6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
 20892-9643, USA
 Tel: 301 443 1706
 Fax: 301 443 9890
 Email: MEST@mail.nih.gov

CDNA library preparation: M.B. Soares Lab Clone distribution:
 Researchers may obtain BMAP CDNA clones from RESEARCH GENETICS. It
 should be noted that Bento Soares is generating a small number of
 additional specialized non-redundant arrays of BMAP CDNAs whose
 availability will be considered under appropriate and limited
 collaborative arrangements. The following repetitive elements were
 found in this CDNA sequence: 1-23, >AT-rich#low-complexity
 Seq primer: M13 Reverse.

FEATURES
 Location/Qualifiers

1..477

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="UI-M-A00-ach-d-01-0-UI"

/clone_1lb="NIH_BMAP_MPG"

/dev_stage="27-32 days"

/lab_host="DH10B (Life Technologies)"

/note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; The
 NIH_BMAP_MPG library is a non-normalized library
 constructed from mouse pineal gland. The tag is a string
 of 5 nucleotides present between the Not I site and the
 oligo-dT track. The library was constructed as described
 by Bonaldo, Lennon and Soares. Genome Research 6: 791-806
 , 1996. Tissue provided by Ms. Annie Novakovich,
 Zivic-Miller Laboratories."

BASE COUNT 149 a 80 c 76 g 172 t
 ORIGIN

Query Match 87.0%; Score 17.4; DB 10; Length 477;
 Best Local Similarity 94.7%; Pred. No. 2.2e+02;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 gtgaatgaccttgcac 19
 |||||

Db 238 GTGAATGCCACTTGCAC 220

RESULT 15
 A4295173

LOCUS
 DEFINITION
 A4295173 610 bp DNA linear GSS 27-JUL-2000
 RPCI-23-132111.TJ RPCI-23 Mus musculus genomic clone RPCI-23-132111
 , DNA sequence.

ACCESSION
 A4295173
 VERSION
 A4295173.1 GI:9536958

KEYWORDS
 GSS.

SOURCE
 house mouse.

ORGANISM
 Mus musculus

REFERENCE
 AUTHORS
 TITLE
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 610)
 Zhao, S., Nierman, W., Feldblum, T., Malek, J., Shatsman, S., Aklnet
 , B., Levins, M., McGann, S., Teagay, G., Geer, K., Krol, M., de Jong, P.
 and Fraser, C.M.

Mouse BAC End Sequences from Library RPCI-23

Unpublished (1999)
 Other_GSSs: RPCI-23-132111.TV

Contact: Shaying Zhao

Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-23. For BAC
 library availability, please contact Pieter de Jong

(pieter@dejong.med.buffalo.edu). Clones may be purchased from

BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>)

or from Resea ch Genetics (http://www.tigr.org/tcb/bac/ends/mouse/bac_end_intro.html)

Plate: 132 row: I column: 11

Seq primer: SP6

Class: BAC ends.

Location/Qualifiers

1..610

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="RPCI-23-132111"

/clone_1lb="RPCI-23"

/sex="Female"

/lab_host="DH10B"

/note="Origin: Kidney/Brain; Vector: pBACe3.6; Site_1:
 EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
 brain genomic DNA was isolated and partially digested
 with a combination of EcoRI and EcoRI Methylase. Size
 selected DNA was cloned into the pBACe3.6 vector at the
 EcoRI sites. The ligation products were transformed into
 DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT 180 a 135 c 94 g 201 t
 ORIGIN

Query Match 87.0%; Score 17.4; DB 12; Length 610;
 Best Local Similarity 94.7%; Pred. No. 2.4e+02;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 2 tgaatgaccttgcact 20
 |||||
 Db 447 TGAATGCCACTTGCAC 465

Search completed: May 22, 2002, 05:31:42
 Job time: 4102 sec

